# Design of OFDM Receivers Operating in Severe Interference Conditions

*An Approach Combining Sparse Signal Reconstruction and Variational Bayesian Inference*

Barbu, Oana-Elena

*DOI (link to publication from Publisher):*
10.5278/vbn.phd.engsci.00094

*Publication date:*
2016

*Document Version*
Publisher's PDF, also known as Version of record

# DESIGN OF OFDM RECEIVERS OPERATING IN SEVERE INTERFERENCE CONDITIONS

AN APPROACH COMBINING SPARSE SIGNAL RECONSTRUCTION AND VARIATIONAL BAYESIAN INFERENCE

**BY**
**OANA-ELENA BARBU**

DISSERTATION SUBMITTED 2016

AALBORG UNIVERSITY
DENMARK

# Design of OFDM Receivers Operating in Severe Interference Conditions

## An Approach Combining Sparse Signal Reconstruction and Variational Bayesian Inference

Ph.D. Dissertation

Oana-Elena Barbu

Dissertation submitted March, 2016

Dissertation submitted: March, 2016

PhD supervisor: Prof. Bernard Henri Fleury
Aalborg University

Assistant PhD supervisor: Assoc. Prof. Carles Navarro Manchón
Aalborg University

Company Supervisors: Tommaso Balercia
Christian Rom
Intel Mobile Communications

PhD committee: Associate Professor Gilberto Berardinelli (chairman)
Aalborg University, Denmark

Specialist Dr. Maja Loncar
Fingerprint Cards, Denmark

Research Director Dr. Pierre Duhamel
Laboratorie de Signaux et Systemes, France

PhD Series: Faculty of Engineering and Science, Aalborg University

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-542-9

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk



Printed in Denmark by Rosendahls, 2016

# Abstract

In this thesis we design iterative signal processing algorithms for OFDM receivers operating in cellular communication systems. The focus is on receiver algorithms capable of coping with harsh propagation and interference conditions, where traditional receiver designs perform poorly. Such conditions include channels with maximum excess delay exceeding the duration of the cyclic prefix, highly Doppler dispersive channels, and severe co-channel interference. These conditions cause intersymbol and/or intercarrier and co-channel interference which severely impair the performance of traditional receivers. We treat these problems from the perspective of designing algorithms for interference-aware receivers operating in 4G-like systems. However, the relevance of our solutions extends beyond the 4G use case as we expect such interference scenarios to become increasingly frequent and important with the deployment of the upcoming 5G systems.

To combat the different types of interference, we advocate that a joint design of the receiver processing that accounts for the interfering signals is required. To realize this joint design, we use tools from compressed sensing and variational Bayesian inference. By exploiting the assumption that the channel impulse response consists of a few non-negligible multipath components, i.e. is sparse, we formulate the channel estimation as a sparse signal reconstruction problem and employ techniques from sparse Bayesian inference to solve it. Then, using variational inference we embed the sparse channel estimator in iterative structures that include the rest of the receiver's tasks: interference cancellation, equalization and data decoding. Numerical evaluations conducted in setups simulating the above harsh conditions show that the receivers implementing the proposed algorithms successfully reconstruct and cancel the interference and perform closely to receivers with perfect interference cancellation capabilities.

To conclude, we highlight the role of interference-aware receiver algorithms in 5G systems. Current and previous cellular systems have adopted a conservative design strategy, in which interference-free reception at the user devices is prioritized at the expense of spectral efficiency. This design has been made with the goal of allowing for low-cost user devices. The increased spectral efficiency requirements and the deployment of dense, heterogeneous cellular

networks expected for 5G systems however, challenge such a design. To meet the future system's targets, it is expected that 5G systems will move part of the computational power demands to the user devices in order to allow for greater flexibility on the network side. With this in mind, advanced receivers implementing algorithms like those developed in this work are potential enablers of a device-centric system architecture.

# Resumé

I denne afhandling designer vi iterative signalbehandlingsalgoritmer til OFDM modtagere i cellulære kommunikationssystemer. Fokus er på modtager algoritmer der kan håndtere barske kanal og interferens betingelser, hvor traditionelle modtagere ikke fungerer. Dette inkluderer trådløse kanaler med maksimal forsinkelse der overstiger varigheden af det cykliske præfiks, kanaler med stor Dopplerspredning og svær ko-kanalinterferens. Sådanne kanaler giver intersymbol og/eller intercarrier og ko-kanalinterferens som alvorligt forringer ydeevnen af traditionelle modtagere. Vi behandler disse problemer ved at designe algoritmer til 4G-lignende systemer, der kan håndtere interferensen. Relevansen af vores løsninger strækker sig ud over 4G systemer, hvor vi forventer sådanne interferens scenarier vil blive hyppigere og vigtigere med introduktionen af kommende 5G-systemer.

For at bekæmpe de forskellige typer af interferens, er vi fortalere for et samlet design af modtageren, der tager højde for de interfererende signaler. For at realisere dette samlede design bruger værktøjer fra komprimeret sensing og variational Bayesiansk inferens. Ved at antage at kanal impuls responset består af nogle få dominerende multipath komponenter, dvs. er sparsom, kan vi formulere kanal estimeringen som en sparsom rekonstruktion og anvende teknikker fra sparsomme Bayesiansk inferens for at løse det. Derefter bruger vi variationel inferens til at integrere den sparsomme kanal estimator i iterative strukturer, der omfatter resten af receiverens opgaver: annullering af interferens, kanal korrektion samt afkodning af data. Numeriske evalueringer der simulerer de ovennævnte barske kanal forhold viser at de foreslåede modtager algoritmer kan rekonstruere og annullere interferensen således de opnår en ydeevne tæt på den der opnås med algoritmer der perfekt kan annullere interferensen.

Afslutningsvis fremhæver vi interferens-opmærksomme modtager algoritmers rolle i 5G-systemer. Nuværende og tidligere cellulære systemer har anvendt en konservativ design strategi, hvor interferens-fri modtagelse på brugerenhederne prioriteres på bekostning af spektral effektivitet. Dette design er blevet anvendt for at tillade billige brugerenheder. De øgede krav til spektral effektivitet og indsættelsen af tætte heterogene cellulære netværk, der forventes i

5G-systemer, anfægter en sådan konstruktion. For at imødekomme de fremtidige systemkrav, forventes det at 5G systemer vil øge beregningskravet til brugerenhederne for at give større fleksibilitet på netværkssiden. Med dette i tankerne, er avancerede modtagere algoritmer, som dem der udvikles i dette arbejde, potentielle katalysatorer for en brugerenheds-centreret system arkitektur.

# Contents

# Thesis Details

| | |
|---|---|
| **Thesis Title:** | Design of OFDM Receivers Operating in Severe Interference Conditions – An Approach Combining Sparse Signal Reconstruction and Variational Bayesian Inference |
| **Ph.D. Student:** | Oana-Elena Barbu |
| **Supervisors:** | Prof. Dr. Sc. Techn. Bernard Henri Fleury<br>Assoc. Prof. Carles Navarro Manchón |
| **Company supervisors:** | Tommaso Balercia<br>Christian Rom |

The main body of this thesis consists of the following papers:

[A] Oana-Elena Barbu, Carles Navarro Manchón, Christian Rom, Bernard H. Fleury, "Message-passing receiver for OFDM systems over highly delay-dispersive channels," *IEEE Transactions on Wireless Communications*, submitted March 2016.

[B] Oana-Elena Barbu, Carles Navarro Manchón, Christian Rom, Tommaso Balercia, Bernard H. Fleury, "OFDM receiver for fast time-varying channels using block-sparse Bayesian learning," *IEEE Transactions on Vehicular Technology*, accepted for publication, February 2016.

[C] Oana-Elena Barbu, Carles Navarro Manchón, Mihai-Alin Badiu, Christian Rom, Tommaso Balercia, Bernard H. Fleury, "Interference-aware OFDM receiver for channels with sparse common supports," *IEEE Global Communications Conference, GLOBECOM*, submitted March 2016.

[D] Oana-Elena Barbu, Niels Lovmand Pedersen, Carles Navarro Manchón, Guillaume Monghal, Christian Rom, Bernard H. Fleury, "Sparse channel estimation including the impact of the transceiver filters with application to OFDM," *15th IEEE International Symposium on Signal Processing Advances in Wireless Communications*, pp. 424 - 428, June 2014.

In addition to the main papers, the following patents have also been granted and/or filed:

[E] Oana-Elena Barbu, Carles Navarro Manchón, Guillaume Monghal, Christian Rom, Bernard Fleury, "Methods and devices for channel estimation and OFDM receiver", *Filed E.U. 2014, Filed U.S.A. 2015.*

[F] Oana-Elena Barbu, Carles Navarro Manchón, Mihai Badiu, Christian Rom, Bernard Fleury, "Methods and devices for channel estimation for mobile systems of insufficient cyclic prefix length", *Filed U.S.A 2015, Filed E.U. 2016.*

[G] Oana-Elena Barbu, Carles Navarro Manchón, Christian Rom, Tommaso Balercia, Bernard Fleury, "Devices and methods for processing one or more received radio signals", *Filed U.S.A. 2015.*

[H] Oana-Elena Barbu, Niels Lovmand Pedersen, Carles Navarro Manchón, Guillaume Monghal, Christian Rom, Bernard Fleury, "Channel estimation technique", *Granted 2015, U.S.A. patent no. US20150003542.*

# Preface

This thesis has been submitted for assessment to the Doctoral School of Engineering and Science in partial fulfillment of the PhD degree. The thesis is based on the published and submitted scientific papers and patents.

The work in this thesis has been carried out under the form of an industrial Ph.D. study at Intel Mobile Communications (IMC) and Aalborg University (AAU) from April 2013 to March 2016. It has been partly financially supported by the Danish Association for Science, Technology and Innovation, Denmark.

At IMC, I have been part of the Verification (VER) unit led by Christian Rom and later on, of the Radio Propagation Modeling (RPM) unit headed by Tommaso Balercia. At AAU, I have been with the Wireless Communication Networks (WCN) section and conducted my research under the supervision of Professor Bernard Henri Fleury and Associate Professor Carles Navarro Manchón. I would like to express my sincere appreciation for the guidance received from my supervisors Bernard, Carles, Christian and Tommaso.

I would also like to express my gratitude to the persons who accepted to be part of my PhD assessment committee, Professor Pierre Duhamel, Specialist Maja Lončar and Associate Professor Gilberto Berardinelli, for taking the time to read and assess this work.

Last but not least, I would like to thank all my colleagues and friends at the WCN section and at IMC.

Oana-Elena Barbu
Aalborg University
April 4, 2016

# Part I

# Introduction

# Chapter 1

# Background and Thesis Overview

In this chapter we discuss the design of the current mobile communication systems and the motivation for deploying the next generation of wireless systems known under the generic name of 5G. In this context, we present the main drivers and requirements for 5G, together with some of the challenges that they will likely impose on the design of the upcoming wireless receivers. We dedicate the second and third sections of this chapter to a presentation of the methodology and the structure of this thesis.

## 1 Brief overview of 5G trends

The design of the future 5G cellular systems is driven by a rapid increase of the number of active users in the network consuming multimedia content. Often, this content has to be delivered to a user equipment operating at high velocities, or at cell-edge. In addition, the number of nodes operating in the network is expected to increase dramatically, since not only human users will be connected. The emergence of machine-type communication will require that the systems allow access to an unprecedentedly high number of devices which inevitably impose big challenges in terms of latency and reliability of the network: while machines connected to the network may not have very stringent bit rate requirements, their operation may be very sensitive to latency and connectivity (e.g. alert and control systems).

Ensuring high data rates and/or low latency in these scenarios is a challenging task for the current 4G cellular systems. Therefore standardization efforts are put towards specifying a set of requirements to enable an enhanced cellular system design, and thus the evolution from 4G to 5G systems. The

new system is expected to deliver data rates of 10+ Gbps and ensure at least 100 Mbps and very low latency (< 1ms) in a network in which the mobile data traffic is doubling every year [1]. These stringent requirements lead to several research initiatives likely to influence the 5G design. The authors of [2] call them "disruptive" trends, emphasizing the idea that only with the redesign of the system's architecture these goals can truly be achieved. We briefly discuss them in the following and point the reader to [2, 3] for a more thorough analysis.

**Millimeter wave communication:** With the exhaustion of the microwave frequencies by the current wireless systems, the community has turned its attention towards millimeter wave (mmW) frequencies, unappealing up to recently due to the severe propagation conditions (high path loss and absorption). However, the deployment of a hybrid architecture consisting of small outdoor cells and isolated indoor cells, and the advancement in antenna techniques make these propagation challenges surmountable. To support this claim, the results of outdoor measurement campaigns [4] showed that the above conditions do not significantly impact the propagation of mmW for cells with radius of $\sim 200$ m, demonstrating the feasibility of using this spectrum in small cells.

**Massive MIMO:** Also known as large scale antenna systems, massive MIMO refers to systems in which the number of antennas at the base station is much larger than the number of users simultaneously allocated to a time-frequency resource [5]. The increased spectral efficiency and the vanishing of noise and interference when the number of antennas grows infinitely large, make massive MIMO an appealing technology. However, its deployment requires not only a full and costly redesign of the hardware structure of the base station, but also enhancing its signal processing capabilities, e.g. the acquisition of the channel state information at the base station becomes challenging as the reuse of pilots across neighboring cells creates pilot contamination.

**Increased densification:** The deployment of uncoordinated small cells, the increase of density of active mobile users in the network, and the emergence of device-to-device communication are all predicted features of the 5G systems. They all however entail the simultaneous usage of time-frequency resources by different devices, and therefore they incur the cost of increased co-channel interference. Combating this type of interference is a highly investigated topic and solutions range from precoding-based interference management at the base station to interference cancellation at the user terminal [6]. However, the heterogeneity of the 5G network and the ad-hoc deployment of the indoor small cells make the interference avoidance task more complicated than ever, requiring flexible solutions both at the base station and at the mobile device.

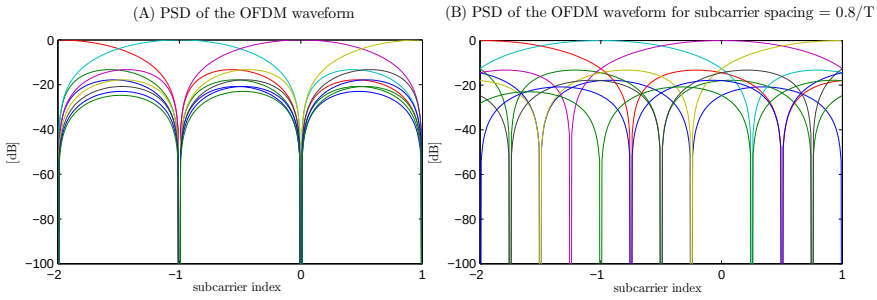


**Fig. 1.1:** Power spectral density (PSD) for an OFDM waveform with subcarrier spacing of (A) $1/T$ and (B) $0.8/T$, also called packed-OFDM, where $T$ is the OFDM symbol duration.

**Advanced receivers:** Also called smart devices, their emergence is favored by the receiver-oriented 5G design and by the fast-paced evolution of the processing capabilities of the modem [7]. These receivers should be able to cope not only with the ripples of network densification but also operate in very high mobility scenarios. Furthermore, given the widespread adoption and well understanding of OFDM for cellular systems, this technology stays in the competition for 5G waveforms [3]. To increase the spectral efficiency, OFDM might suffer changes such as shortening/discarding the cyclic prefix, reducing the subcarrier spacing (see the example in Fig. 1.1 and [3] for a more detailed discussion), etc. making the receiver prone to intercarrier and intersymbol interference. All these conditions challenge the operation of the OFDM receiver and require the redesign of the algorithms implemented in it.

While the technologies listed above are expected to impact the deployment of 5G systems, combining them in a successful system design will undoubtedly be a challenging task, both technologically and economically, requiring the cooperation of standardization, academic and industry bodies.

# 2 Problem statement and research methodology

In this work, we are focusing on the challenges that the severe interference conditions pose on the functioning of the wireless receiver. Namely, we investigate the operation under self-interference, due to delay- and Doppler- dispersion in the channel, and co-channel interference. These conditions already pose problems to the operation of 4G modems and, as discussed in this chapter, they will be drivers for the evolution towards 5G systems. Furthermore, although yet to be defined, the 5G systems are likely to bear some resemblance to 4G, by utiliz-

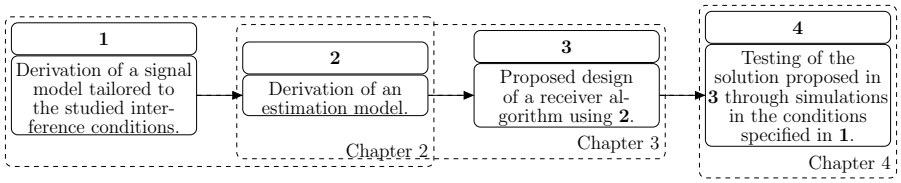


**Fig. 1.2:** Research methodology

ing similar technologies, e.g. OFDM-like transmission. Finally, encouraged by the evolution towards advanced receivers, we design receiver algorithms able to cope with these harsh conditions by relying on study cases of 4G-like systems. Choosing these study cases allows us to clearly formulate the problem, design appropriate solutions, test them and draw conclusions which will allow us in the future to devise receiver solutions in the 5G setup – once all its "ingredients" have been specified.

We apply the methodology described in Fig. 1.2 to conduct our studies of receiver design for the above conditions: each study consists of deriving a signal model and a subsequent approximate model which enable us to cast a simple probabilistic model. Based on the latter we design receiver algorithms which we finally test through simulations using the initial signal model. This methodology can be recognized in each of the articles (Papers A-D) that conform the main contribution of the thesis. In addition, the remaining chapters of Part I follow this methodology by offering insights into: the formulation of each studied problem (Chapter 2), the mathematical tools used in the design of the solutions (Chapter 3), and the conclusions of our studies (Chapter 4).

Lastly, this project is a joint endeavor meant to facilitate the knowledge exchange between industry and academia. The study of the above harsh interference conditions and the consequential algorithm development enable us to formulate conclusions that the chip manufacturers can use as guidelines to revise the design of the baseband modems for the current cellular systems and develop an enhanced design for the future generations of modems.

# 3 Thesis structure

The remainder of this thesis is organized as follows.

**Part I** consists of Chapter 1 (this chapter) to Chapter 4.

**Chapter 2** analyzes three main causes of interference studied in this thesis. The analysis is made in the context of 4G with an outlook towards 5G systems,

the deployment of which is expected to make the interference problem even more severe. Each section of the chapter discusses one of the three causes: Section 1 looks at the operation of the receiver in OFDM systems with reduced (or discarded) CP. Section 2 deals with the transmission over highly Doppler dispersive channels and Section 3 treats the co-channel interference problem arising from the synchronous usage of the same time-frequency resources by several users. Lastly, Section 4 introduces some strategies for dealing with interference by going beyond the conventional capabilities of current receivers.

**Chapter 3** discusses the mathematical tools used in the design of solutions to the problems enunciated in Chapter 2. Section 1 introduces variational Bayesian inference for solving joint optimization problems and presents the mean-field belief-propagation framework that we use in the design of the receiver algorithms. Section 2 presents the sparse Bayesian inference methodology and background. The chapter ends with examples of how these tools have been applied in the field of wireless communications.

**Chapter 4** lists the main contributions of this thesis and summarizes the conclusions of the study.

**Part II** includes the papers A to D the contributions of which have been listed in Chapter 4.

In this chapter we introduced the main 5G trends and challenges. Given the broadness of this topic and the ongoing standardization efforts, we limited the discussion to general concepts that are believed to influence the deployment of the new system and discussed the impact of the interference on the operation of the receiver. We will retake in more detail the interference analysis, e.g. types, causes, and discuss solutions to mitigate it in the next chapters following the structure presented above.

# Chapter 2

# Severe Interference in OFDM: an Outlook Towards 5G

In this chapter we bring forward three interference scenarios which encumber the operation of the wireless OFDM receiver. Already relevant in the 4G context, these scenarios represent particularly important challenges for 5G systems, the deployment of which will make them frequent use cases. In the following, we discuss how interference impairs the receiver performance in these scenarios, and what design changes the receiver algorithm needs to undergo in order to overcome the performance degradation.

In Section 1 we discuss the self-interference arising as a result of utilizing a cyclic prefix (CP) of insufficient duration. In Section 2 we tackle the intercarrier interference problem arising when operating over highly Doppler-dispersive channels, while in Section 3 we discuss co-channel interference as a result of network densification.

## 1 Towards a shorter cyclic prefix

The wireless CP-OFDM transmitter generates OFDM waveforms including a cyclic prefix (CP) from a transmit bit stream. As long as the duration of the CP is larger than the maximum excess delay of the multipath wireless channel, the transmission is interference free. The insertion of CP is however a compromise solution since it results in spectral efficiency loss which is directly proportional to the CP duration. Two characteristics of the OFDM waveform are preserved at reception: (i) the subcarriers remain orthogonal to each other and (ii) no previous OFDM waveforms overlap with the current one. These characteristics

are fundamental to the operation of the classical wireless receiver whose entire design is based on the assumption that (i) and (ii) are always true. What (i) and (ii) imply is that the observed signal at any subcarrier is the transmit signal at the same subcarrier weighted by the corresponding channel gain and corrupted by noise. The resulting structure of the observed signal is one of the advantages of the OFDM transmission and enables the receiver to implement a simple and fast equalization scheme, and to operate on the received signal in a sequential manner: estimate channel responses using apriori known symbols (pilots), equalize, demodulate and decode. This is an appealing property of the technology, and together with its widespread usage in wireless networks is a strong argument that favors the adoption of OFDM in 5G.

However, since the spectrum has become an expensive commodity, OFDM spectral efficiency could be improved by relaxing the orthogonality requirements and/or by shortening or even removing the CP. Such a departure from OFDM is made at the expense of introducing intercarrier and intersymbol interference (generically called self-interference). Nevertheless, allowing for some amount of self-interference and properly equipping the receiver to deal with it, might prove to be a beneficial strategy. In fact, the wireless community is already advocating the design of so-called *advanced receivers* endowed with higher computational power and that implement algorithms capable of more complex equalization and decoding [2, 7, 8].

Following the above reasoning, one part of our work studies the design of receiver algorithms that cope with self-interference arising from CP of insufficient duration in OFDM systems. Two arguments motivate our choice: (a) this situation, although not very frequently, does occur in currently deployed 4G systems, and (b) since the problem is likely to become increasingly relevant in a 5G context, the conclusions of our study could provide useful insights into the design of the future receivers. Therefore, having laid out the setting of this study, we pose the following questions:

- How can accurate channel state information still be acquired when the pilots are also corrupted by intercarrier and intersymbol interference?
- How can accurate equalization of a received signal corrupted by self-interference be performed?

The straightforward answer to the two questions above is that *the presence of self-interference calls for a rethinking of the classical receiver design.* Channel estimation, equalization and decoding should be treated as a joint task by an interference aware receiver, which means that channel responses and data estimates need to be refined iteratively. However, since highly inaccurate estimates are expected to be computed in the first iterations, to really benefit from such a structure, a receiver algorithm should account for the uncertainties in these estimates. This can be accomplished by using soft values for both the channel and data estimates. Although rarely used in the context of self-interfering

OFDM transmission due to insufficient CP length, the idea of operating with soft data estimates to iteratively cancel the interference has been previously explored [9, 10]. Harnessing this idea however was made without changing the underlying design of the receiver, traditionally based on breaking down the task of recovering the transmitted information bits into smaller independent subtasks such as channel estimation, equalization and decoding. Hence, such designs rely on the assumption that accurate channel state information is available at the receiver prior to equalization. As we have discussed above, this assumption is completely unrealistic since the pilots are corrupted by interference and they do not provide a dense enough sampling to capture the strongly selective behavior of the channel frequency response. In this work we move away from such a design and formulate a joint optimization problem employing the tools described in Chapter 3.

# 2 Towards receivers for high Doppler scenarios

In this section we discuss the operation of the wireless receiver in highly Doppler-dispersive environments. The motivation for this study is two-fold:

**i)** 4G/LTE systems have been designed to support mobility but not to simultaneously optimize for performance. Namely, the design of the OFDM receiver relies on the assumption that strict subcarrier orthogonality is preserved, and coping with scenarios in which intercarrier interference severely affects the quality of the reception is accomplished through retransmissions, adapting the modulation and coding scheme etc. Aiming for low complexity receivers, 4G system design promotes trading off spectral efficiency for reliability. However, with the expected increase in computational power in the mobile devices, an interference-aware receiver, able to cancel or reject the interference from the received signal, might represent a viable compromise solution for the future.

**ii)** In 5G systems, solutions for coping with high-Doppler are much more relevant due to the emergence of two trends: 1) operation at mmW frequencies and 2) high performance for high velocity users. On the one hand, the scarcity of the microwave spectrum makes the utilization of the frequencies in the 3 – 300 GHz range inevitable for the future 5G systems [2]. On the other hand, 5G systems need to ensure a good receiver performance in very high mobility scenarios, e.g. car or train to infrastructure communication, which are increasingly common [11]. It is important to note that 1) and 2) do not occur simultaneously, since the usage of mmW is reserved for small cells accommodating users with low or moderate mobility, while high speed users will be served by base stations operating in microwave frequency bands. Nevertheless,

each trend considered individually forces the receiver to operate over highly Doppler dispersive channels.

In this work, we study the design of receivers operating over high-Doppler channels in 4G-like systems. We expect that the obtained conclusions will be useful to the design of both the waveforms and parameters of 5G systems, and their corresponding receivers. The classical OFDM receiver operating over such channels experience intercarrier interference, a condition for which it has not been designed. Since the interference affects both data and pilot subcarriers, the performance of both the channel estimator and the decoder will be affected, and a similar situation as that discussed in Section 1 arises. During one OFDM symbol, the receiver observes at subcarrier $k$ a signal $y_k$ which is the superimposition of the transmit symbol $x_p$ at every subcarrier, indexed by $p$, each weighted by a channel gain $H_{k,p}$ and corrupted by AWGN $w_k$: $y_k = H_{k,k}x_k + \sum_{p \neq k} H_{k,p}x_p + w_k$. At high Doppler frequencies, the rapid time-varying channel impulse response incurs the loss of subcarrier orthogonality and thus a power leakage among all subcarriers. A visual representation of this effect is given in Fig. 2.1.

Historically, canceling intercarrier interference has been treated as an equalization problem and solutions have been designed under the assumption that accurate channel state information is available [12–15]. In a system in which pilot symbols are multiplexed in frequency with data symbols this assumption is not justified. Our solution relies on the design of algorithms which account for the Doppler effect both when acquiring the channel state information and when equalizing the signal [16, 17]. In our work we follow this approach and find answers to the following questions:

- What channel features can a receiver algorithm exploit to reconstruct the time variations of the channel response?
- What is the performance improvement of an iterative scheme that uses soft symbol estimates to refine the channel estimates and vice versa?

# 3 Coping with network densification

The emergence of 5G systems favors a heterogeneous network architecture, in which ad-hoc densely deployed small cells, that share the spectrum with each other and with the overlay network create co-channel interference [1, 18]. The network heterogeneity makes the interference much harder to mitigate using network-level interference management schemes as previous cellular systems do. This is only one of the 5G scenarios which favors the occurrence of co-channel interference. Another one is the increased number of user equipments simultaneously active that operate in the vicinity of each other, but belong to different networks (i.e. are served by base stations which do not cooperate). Ei-

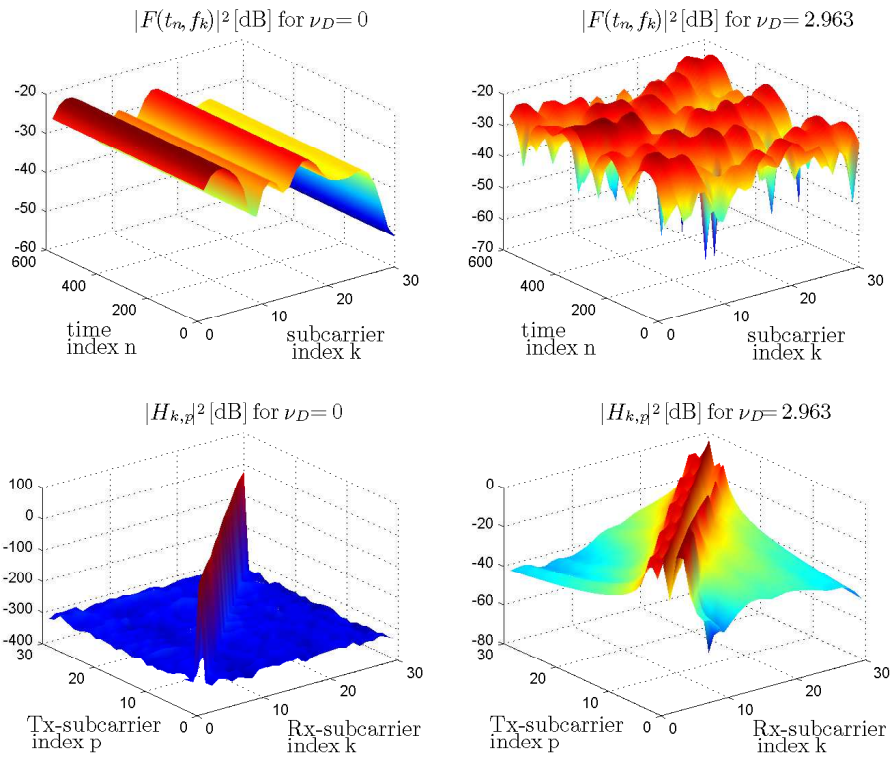


**Fig. 2.1:** Sampled channel transfer function $F(t_n, f_k)$ (top left and right), and entries of the channel frequency matrix $\mathbf{H}$ (bottom left and right), corresponding to one realization of the channel impulse response $h(t,\tau) = \sum_{l=1}^{L} h_l \exp(2\pi j f_l t)\delta(\tau - \tau_l)$ for different maximum Doppler frequencies. The time-varying channel impulse response consists of $L$ multipath components, where $h_l$, $f_l = \nu_D f_s \cos\theta_l$ and $\tau_l$ are respectively the complex weight, Doppler shift and delay of the $l$th multipath component, $T_s$ is the sampling time of the system, $f_s$ is the subcarrier spacing and $\nu_D$ is the maximum Doppler shift normalized to $f_s$, $t_n = nT_s, f_k = kf_s$.

ther scenarios imply a large number of wireless devices causing to the so-called phenomenon of "network densification". Studies of this phenomenon and its implications gained popularity in recent years and the tremendous number of research articles dealing with co-channel interference stands proof of that. Since it became a thoroughly investigated research topic, diverse solutions to cope with interference have been proposed and they range from system redesign to distributed processing among mutually interfered devices, see [19–21] and the references therein. One can claim that a common denominator of these solutions is the 5G-favored "device centric architecture" supported by the development of smart devices of high computational power i.e. the advanced receivers [2].

Given the broadness of this topic, we restrain the discussion to solutions that challenge the redesign of the classical wireless MIMO OFDM receiver. We further narrow down the scope to consider only algorithms that cancel interference locally without cooperating with the interferers. We study the scenario in which several devices are synchronously receiving data in the same frequency bands and given their proximity, they interfere on each other. In this case, the receiver, being interference aware, has two main options: either to neglect the interference or to attempt at estimating and canceling it. In this work, we are seeking answers to the following research questions:

- When estimating the MIMO sub-channels, of either desired or interfering users, which are the channel features that can be leveraged to design the channel estimator?
- When the receiver has access to the modulation alphabets and codes of the interferers, to what extent does it benefit from exploiting this information (e.g. to estimate the interfering signals). Otherwise, what alternative approaches can the receiver undertake to estimate and cancel co-channel interference?

# 4 Receiver design following the 5G trends

The take-home message from the presentation of the development trends towards 5G from Sections 1, 2, and 3 is that interference (be it self-induced or of multiuser nature) will be increasingly common in the future wireless communication systems and will impose important challenges on the design of the 5G receiver. This receiver has to acquire accurate channel information and accurately estimate the desired bit stream using signals corrupted by interference. Since the two tasks are interdependent, there is a need for a design that accounts for interference both when performing channel estimation and detection. In the following we introduce two approaches which enable such a

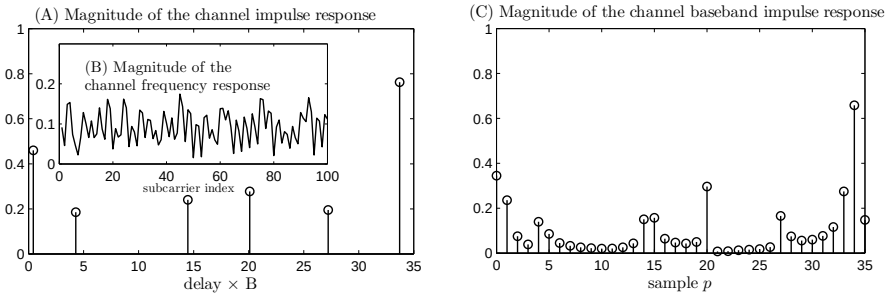


**Fig. 2.2:** Magnitudes of (A) a realization of a sparse channel impulse response $h(\tau) = \sum_{l=1}^{L} h_l \delta(\tau - \tau_l)$, where $h_l$ and $\tau_l$ are the gain and delay of the $l$th multipath component, (B) the channel transfer function $F(kf_s) = 1/\sqrt{N} \sum_{l=1}^{L} h_l \exp(-2\pi j k f_s \tau_l)$ and (C) the channel baseband impulse response $\alpha_p = 1/\sqrt{N} \sum_{k=0}^{N-1} F(kf_s) \exp(2\pi j k p/N), p \in \{0, 1, ..., N-1\}$ for a bandwidth $B = Nf_s$, with $N$ subcarriers spaced by $f_s$.

design. In the next chapter, we will introduce the theoretical foundations of the methods used for this design.

### Application of variational Bayesian inference to design receiver algorithms

The deployment of 5G mandates the device-centric architecture of a network consisting of advanced receivers. This context favors the design of more complex algorithms for interference-aware receivers. In this work we resort to variational Bayesian inference to design algorithms able to cope with the harsh interference conditions. The proposed receiver algorithms are tailored to OFDM systems experiencing interference arising in each of the scenarios described in Sections 1, 2, and 3. Their common denominator is the joint design of all receiver's tasks. The design is enabled by casting an approximate system model and associating to it a probabilistic model. Then, applying variational techniques allow us to estimate all the unknowns of the system (channel responses, noise variance, information bits from the desired and interfering users, etc.).

### Application of compressed sensing to design channel estimators

This approach is motivated by a particular feature of the wireless propagation channel. In many propagation environments, the impulse response of the wireless propagation channel can be characterized as the superimposition of a few multipath components of non-negligible power, each with its own complex gain and delay, i.e. the channel is sparse [22–24]. The sparsity property of the channel translates in the fact that the channel impulse response consists of a

number of components much smaller than its baseband representation [25] – see Fig.2.2.

Delay-domain channel sparsity manifests itself under different forms: when the propagation environment consists of big obstacles, the wireless channel exhibits a few clusters of multipath components in delay domain and the literature refers to such channel as *clustered-sparse* [26]; MIMO subchannels exhibit *sparse common support* [27] meaning that the delay supports of the impulse responses of the sub-channels at different receive antennas are identical due to the small antenna separation. Traditional OFDM channel estimation does not take advantage of this property and employs various training based methods [28, 29] to retrieve the channel frequency responses which are used afterward for equalization.

Becoming increasingly popular, the strategy of exploiting the channel sparsity proves to be advantageous for a receiver operating in various "harsh" interference scenarios, when the pilot signals alone are not sufficient to perform accurate channel estimation. As a result, more and more channel estimation schemes successfully apply techniques from compressed sensing by capitalizing on the delay-domain sparsity [22, 23, 26, 27, 30].

# Chapter 3

# Bayesian Inference Frameworks

In this chapter we introduce the mathematical tools used in the design of the proposed receiver algorithms. The task of the receiver algorithm is to recover the transmitted bits vector $\mathbf{u} = [u_1, .., u_m, ..., u_N]$ from a received signal vector $\mathbf{y} = f(\mathbf{H}, \mathbf{u}, \mathbf{w})$, where $f(\mathbf{H}, \mathbf{u}, \mathbf{w})$ stands for all the transformations that the bit vector $\mathbf{u}$ undergoes before it reaches the receiver, with $\mathbf{H}$ and $\mathbf{w}$ denoting the channel matrix and noise vector respectively. To minimize the bit error probability, the MAP bit decoder is employed, i.e. $\hat{u}_m = \arg\max_{u_m} p(u_m|\mathbf{y})$. Solving the MAP problem is however typically intractable since neither $\mathbf{H}$ nor $\mathbf{w}$ are known, hence, classical receiver algorithms split the problem of recovering $\mathbf{u}$ from $\mathbf{y}$ into simpler problems. A typical approach consists of dividing the receiver functionality into three tasks: channel estimation, equalization and decoding. The tasks are performed sequentially by different receiver's blocks.

By contrast, in this work, we aim at designing a receiver algorithm in a holistic manner: we use the variational inference framework to formulate the problem of recovering the transmitted bits as an optimization problem with a global objective function. Specifically, we use a method combining mean-field (MF) and belief propagation (BP), as it has favorable properties that will be detailed in Section 1. In Section 2 we turn our attention to sparse Bayesian inference (SBI) which we employ to design sparse estimators of the channel impulse response. The Bayesian formulation of MF-BP and SBI makes them easily incorporable in the joint receiver design.

# 1 Variational Bayesian inference

This section is organized as follows: Section 1.1 introduces concepts from variational Bayesian inference, Section 1.2 explains the factor graph representation and Section 1.3 describes the MF-BP algorithm in its message-passing implementation.

## 1.1 Brief introduction

Variational Bayesian inference (VBI) encompasses tools/methods for computing approximations of the marginal probability density functions (pdfs)/pmfs, of latent variables in a probabilistic system. Here, the word *variational* is broadly used to designate "mathematical tools that express the quantity of interest as the solution of an optimization problem." The optimization problem is often relaxed "by approximating the function to be optimized [...]. Such relaxations, in turn, provide a means of approximating the original quantity of interest " [31]. They aim at replacing the original intractable optimization problem by a feasible one.

Consider a probabilistic system consisting of a set $\mathcal{Z}$ of unknown (latent) random variables and characterized by a joint pdf that factorizes according to

$$p(\mathbf{z}) \propto \prod_{f_a \in \mathcal{F}} f_a(\mathbf{z}_a), \tag{3.1}$$

where $\mathbf{z}$ is a vector containing all random variables $z_i \in \mathcal{Z}$, $\mathbf{z}_a$ is the vector of all variables $z_i$ that are argument of the local function $f_a$ included in the set of functions $\mathcal{F}$. In our application, $p(\mathbf{z})$ is the posterior distribution of the latent variables in the probabilistic system model given an observation $\mathbf{y}$. Applying Bayesian inference on the model (3.1) is about computing some marginals of this joint pdf. In complex systems (i.e. with $|\mathcal{Z}|$ very large and complex interdependencies between the elements of this set) the computation of these marginals is not always tractable, so we have to resort to efficient, feasible methods to solve this problem.

The VBI framework provides iterative methods that compute so-called beliefs, which - at best - coincide with the marginals in special cases, and otherwise are approximations thereof. In this framework the inference problem is formulated as that of minimizing the Gibbs free energy, or an approximation of it, [32] for the model (3.1). Solving the minimization problem for a selected free energy leads to a set of stationary point equations that are implicit functions of the beliefs. The sought iterative algorithm is a fixed-point algorithm derived from the stationary point equations[1]. In this work we make use of three fixed-point algorithms: MF [33], BP [34], and a combination of both called

[1]Many fixed-point algorithms can be derived from a set of implicit equations.

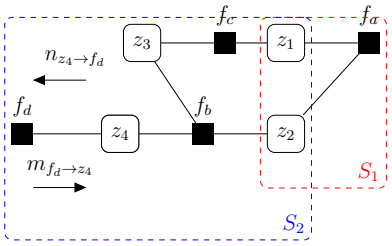


**Fig. 3.1:** Example of factor graph representation of $p(z_1, z_2, z_3, z_4) = f_a(z_1, z_2) f_b(z_2, z_3, z_4) f_c(z_1, z_3) f_d(z_4)$. $S_1$ and $S_2$ are examples of two disjoint subgraphs.

MF-BP [35]. These iterative algorithms are described in their message-passing implementation in Section 1.3.

## 1.2 Factor graph representation

The (bipartite) factor graph of a probabilistic model $p(\mathbf{z})$ provides a visual representation of the factorization in (3.1). It consists of a variable node for each variable $z_i$, a factor node for each factor $f_a$, and edges which connect a variable node $z_i$ to a factor node $f_a$ [2] if, and only if, $z_i$ is an argument of $f_a$. Similarly, a subgraph is a bipartite graph that includes a subset of the factor nodes of the graph, all nodes of the variables connected to the factors of this subset, and the corresponding edges from factor to variable nodes. We define two subgraphs of a bipartite factor graph to be disjoint if the subsets of factors in the subgraphs are disjoint. Note that two disjoint subgraphs can have common variable nodes. We use $\mathcal{N}(z_i) \subseteq \mathcal{F}$ to denote the subset of functions $f_a$ that have variable $z_i$ as an argument and $\mathcal{N}(f_a) \subseteq \mathcal{Z}$ for the subset of variables that are arguments of $f_a$ – see the example in Fig. 3.1.

VBI methods we are interested in can be formulated as message-passing algorithms that iteratively exchange messages along the edges of the factor graph. For each edge, two messages are computed: one from the variable node to the factor node and vice-versa, where by *message* we understand a positive real-valued function of the variable associated with the variable node. The message from variable node $z_i$ to the neighboring factor node $f_a$ is denoted by $n_{z_i \to f_a}$ and the message from $f_a$ to $z_i$ is $m_{f_a \to z_i}$. A message-passing algorithm typically consists of an initialization phase, in which some messages in the graph are initialized and an iterative phase, in which messages are sequentially updated according to a specified scheduling until convergence or a stopping criterion is satisfied. The beliefs are then retrieved from the values of the

[2] By abuse of language we identify a variable (factor) node with the variable (factor) it represents.

messages.

## 1.3 Message passing formulation of the combined MF-BP

In the following we briefly introduce the MF and BP methods and present the message-passing approach resulting from combining them.

MF iteratively computes a "simplified" belief of the joint pdf in (3.1) that minimizes the Gibbs free energy. The tractability of the minimization is ensured by choosing a simple structure for this belief, e.g. a full factorization in the beliefs of single latent variables, or a partial factorization in beliefs of groups of said variables. The two variants are called naive MF and structured MF respectively [36]. MF algorithms yield simple update expressions for conjugate exponential models, guaranteeing convergence (by ensuring that the variational free energy is non-increasing from one iteration to the other) but, in general, not the exact solution. Another drawback of the method is that it is not compatible with hard constraints [35].

The fixed points of BP are the stationary points of the Bethe free energy [37] for the model (3.1). BP applied on a factor graph computes exact marginals of the global pdf in (3.1) if the factor graph has a tree structure. The compatibility of the BP with hard constraints and the computation of exact marginals in a tree graph make the method a very appealing choice. If the factor graph contains cycles however, the beliefs yielded by BP are only approximations of the marginals. Note that the factor graphs of the probabilistic systems considered in this work precisely share this property. Lastly, applying BP to a probabilistic model with both discrete and continuous variables may generate highly complex algorithms.

As discussed above, MF and BP have their own pluses and minuses. Hence, a solution which combines them, and thus exploits the virtues of both and circumvents their drawbacks, is highly beneficial. Such a solution is developed by the authors of [35] who propose a unified message-passing approach which combines the two methods. The method consists of decomposing the factor graph into two disjoint subgraphs, each implementing either MF or BP, and specifies how to propagate messages between the two subgraphs. The fixed points of MF-BP are the stationary points of a region-based free energy for the model (3.1) [35]. Both MF and BP can be derived as particular instances of MF-BP.

Specifically, we consider the joint pdf (3.1) and we define two subgraphs, a MF subgraph and a BP subgraph, which together contain all the factor nodes in the graph. We use $\mathcal{F}_{\mathrm{MF}}(\mathcal{F}_{\mathrm{BP}})$ to denote the set of factor nodes in the MF (BP) subgraph, where $\mathcal{F}_{\mathrm{MF}} \cap \mathcal{F}_{\mathrm{BP}} = \emptyset$ and $\mathcal{F}_{\mathrm{MF}} \cup \mathcal{F}_{\mathrm{BP}} = \mathcal{F}$. Given this decomposition, the joint pdf (3.1) can be expressed as

$$p(\mathbf{z}) \propto \prod_{f_a \in \mathcal{F}_{\mathrm{MF}}} f_a(\mathbf{z}_a) \prod_{f_b \in \mathcal{F}_{\mathrm{BP}}} f_b(\mathbf{z}_b) \tag{3.2}$$

The method in [35] specifies the messages to be passed in each subgraph and at the variable nodes belonging to both subgraphs. The message from a factor node $f_a \in \mathcal{F}_{\text{MF}}$ to the variable node $z_i \in \mathcal{N}(f_a)$ reads

$$m^{\text{MF}}_{f_a \to z_i}(z_i) \propto \exp\left( \sum_{\mathbf{z}_a \setminus z_i} \log f_a(\mathbf{z}_a) \prod_{j : z_j \in \mathcal{N}(f_a) \setminus z_i} n_{z_j \to f_a}(z_j) \right). \tag{3.3}$$

The message from a factor node $f_a \in \mathcal{F}_{\text{BP}}$ to the variable node $z_j \in \mathcal{N}(f_a)$ is

$$m^{\text{BP}}_{f_a \to z_i}(z_i) \propto \sum_{\mathbf{z}_a \setminus z_i} f_a(\mathbf{z}_a) \prod_{j : z_j \in \mathcal{N}(f_a) \setminus z_i} n_{z_j \to f_a}(z_j). \tag{3.4}$$

We note that the update rule for the messages outgoing factor nodes belonging to the MF (BP) subgraphs coincides with the standard MF (BP) rule for such messages. The message passed by a variable node $z_i \in \mathcal{Z}$ to a factor node $f_a \in \mathcal{N}(z_i)$ is

$$n_{z_i \to f_a}(z_i) \propto \prod_{f_c \in \mathcal{N}(z_i) \cap \mathcal{F}_{\text{MF}}} m^{\text{MF}}_{f_c \to z_i}(z_i) \prod_{f_c \in \mathcal{N}(z_i) \cap \mathcal{F}_{\text{BP}} \setminus f_a} m^{\text{BP}}_{f_c \to z_i}(z_i). \tag{3.5}$$

Lastly, to retrieve the belief of $z_i \in \mathcal{Z}$ the following rule is applied:

$$q(z_i) \propto \prod_{f_c \in \mathcal{N}(z_i) \cap \mathcal{F}_{\text{MF}}} m^{\text{MF}}_{f_c \to z_i}(z_i) \prod_{f_c \in \mathcal{N}(z_i) \cap \mathcal{F}_{\text{BP}}} m^{\text{BP}}_{f_c \to z_i}(z_i). \tag{3.6}$$

Note that by setting $\mathcal{F}_{\text{MF}} = \emptyset$ ($\mathcal{F}_{\text{BP}} = \emptyset$) we obtain the computation rules for BP(MF).

# 2 Sparse Bayesian inference

Sparse Bayesian inference (SBI) is a generic framework that encompasses methods for reconstructing sparse signals from a noisy linear system. This is accomplished by formulating a fully probabilistic model of the system and imposing a prior distribution on the signals of interest which favors sparse solutions (a so-called sparsity inducing prior) [38, p.212]. Reconstruction algorithms belonging to SBI are generally invoked when the system is underdetermined (thus with many solutions) and additional characteristics of the signal, such as sparsity, need to be exploited [39]. Imposing sparsity-inducing priors can therefore be thought of as a method that implements Occam's razor. In the following we describe SBI for reconstructing sparse signals in the single and multiple measurement vector models.

## 2.1 Single measurement vector (SMV) model

Sparse signal recovery [38, 40] deals with recovering a signal $\mathbf{t} \in \mathbb{C}^N$ that admits a sparse representation in the following model:

$$\begin{aligned} \mathbf{y} &= \mathbf{t} + \mathbf{w} \\ \mathbf{t} &= \mathbf{\Phi}\mathbf{x} \end{aligned} \tag{3.7}$$

where the matrix $\mathbf{\Phi} \in \mathbb{C}^{N \times M}$ is called a dictionary matrix, and $\mathbf{w}$ is a white Gaussian noise vector with precision (inverse variance) $\lambda$. The signal $\mathbf{t}$ admits a $K-$sparse representation if it can be expressed as the linear combination of $K \ll M$ columns of $\mathbf{\Phi}$, i.e. $\mathbf{t} = \sum_{k \in \mathcal{K}} \boldsymbol{\phi}_k x_k$, where $\boldsymbol{\phi}_k$ is the $k$th column of $\mathbf{\Phi}$, and the set $\mathcal{K}$ of cardinality $K$ contains all the indices of the non-zero entries of $\mathbf{x}$. This implies that the vector $\mathbf{x} \in \mathbb{C}^M$ has only $K$ non-zero entries, i.e. $\mathbf{x}$ is sparse. The goal becomes to estimate the entries of the sparse vector $\mathbf{x}$ given the observation $\mathbf{y} \in \mathbb{C}^N$ and the dictionary $\mathbf{\Phi}$.

Based on the above considerations, one approach is to formulate the MAP estimation problem: $\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ where $p(\mathbf{y}|\mathbf{x}) = \mathrm{CN}(\mathbf{y}|\mathbf{\Phi}\mathbf{x}, \lambda^{-1}\mathbf{I}_N)$, or, equivalently, $\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y}) = \arg\min_{\mathbf{x}} \| \mathbf{y} - \mathbf{\Phi}\mathbf{x} \|_2^2 + \lambda^{-1} g(\mathbf{x})$, where $g(\mathbf{x}) \propto -\log(p(\mathbf{x}))$. [3] This problem is a regularized least-squares regression with regularization term $g(\mathbf{x})$ dependent on the choice of prior $p(\mathbf{x})$. The aim is thus to compute $\hat{\mathbf{x}}$ by choosing a computationally convenient prior $p(\mathbf{x})$ which favors sparse solutions. Common choices are the Laplacian prior, [41], Jeffreys prior [42, 43], etc. Note that the choice of prior may not reflect the true pdf of $\mathbf{x}$. The prior is selected in such a way that the resulting MAP estimator promotes sparsity.

An alternative approach to estimating sparse $\mathbf{x}$ in (3.7) is to select a hierarchical prior model for $\mathbf{x}$, i.e. a prior dependent on some latent variables $\boldsymbol{\gamma}$ that should also be estimated from the available observations $\mathbf{y}$. With the appropriate choice of $p(\boldsymbol{\gamma})$, priors such as Laplacian, Jeffreys etc. can be expressed using a hierarchical prior model. The optimization problem is then reformulated as MAP estimation of $\boldsymbol{\gamma}$. Once an estimate $\hat{\boldsymbol{\gamma}}$ is obtained, the mean of the posterior $p(\mathbf{x}|\hat{\boldsymbol{\gamma}}, \mathbf{y})$ is a "commonly accepted point estimate" [44] of $\mathbf{x}$.

Note that the two approaches above have been termed Type I and Type II estimation, respectively. For a more detailed discussion on the choice of priors and its implications on the performance of the estimators employing them we refer the reader to [44].

Hierarchical models have been successfully applied to the design of sparse channel estimators. We focus the discussion on two of such sparsity inducing priors that have been used in our work: the Gamma-Gaussian prior [38, 45, 46]

[3] Here $x \propto y$ denotes $x = ay$ for some positive constant $a$.

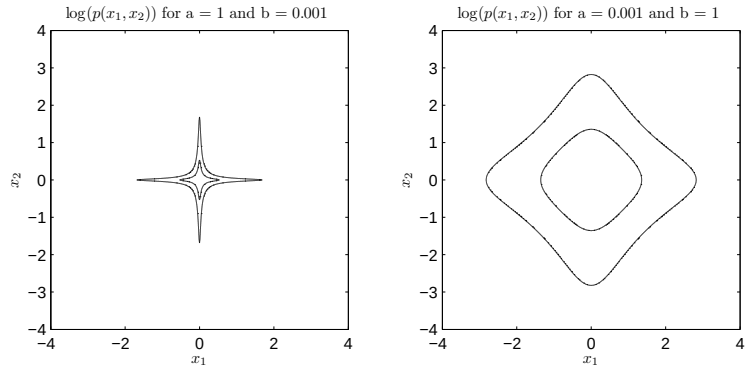


**Fig. 3.2:** Contour plots of the restriction to $\mathbb{R}^2$ of $\log(p(x_1, x_2))$ for two settings of $a$ and $b$.

and the Bernoulli-Gaussian prior [47, 48]. [4]

The Gamma-Gaussian hierarchical modeling of $\mathbf{x}$ is

$$\begin{aligned} p(\mathbf{x}, \boldsymbol{\gamma}) &= \prod_{m=1}^{M} p(x_m|\gamma_m)p(\gamma_m) \\ p(x_m|\gamma_m) &= \mathrm{CN}(x_m|0, \gamma_m^{-1}) \\ p(\gamma_m) &= \mathrm{Gam}(\gamma_m|a, b) \end{aligned} \tag{3.8}$$

where $\mathrm{CN}(\cdot|\mu, \sigma^2)$ is the complex Gaussian pdf with mean $\mu$ and variance $\sigma^2$, $\mathrm{Gam}(\cdot|a, b)$, is the Gamma pdf with shape and rate parameters $a$ and $b$ respectively, and $\gamma_m$ is the precision of the $m$th weight. Selecting large values for $\gamma_m$ (e.g. by setting a large value for the shape $a$ or a small one for the rate $b$) leads to $p(x_m|\gamma_m)$ being very peaked around zero.

We illustrate the sparsity-inducing property of this prior in the following. With the model (3.8), we obtain the marginal $p(\mathbf{x}) = \int p(\mathbf{x}, \boldsymbol{\gamma}) d\boldsymbol{\gamma} = \prod_{m=1}^{M} p(x_m)$ where $p(x_m) \propto (|x_m|^2 + b)^{-(a+1)}$. In Fig 3.2 we depict the contours of the restriction of $p(\mathbf{x})$ to $\mathbb{R}^M$ for $M = 2$ and different values of $a$ and $b$. We observe that the larger $a$ grows or conversely, the smaller $b$ becomes the more concentrated along the $\mathbf{x}-$axes $p(\mathbf{x})$ becomes. This concentration along the axes promotes sparsity as shown in the numerical studies presented in [46].

In the Bernoulli-Gaussian model, each sample $x_m$ is described as $x_m =$

[4] Although in this work Type II estimators are used, we restrict ourselves to discussing the sparsity-inducing properties of the hierarchical priors from the Type I estimation perspective. This choice is motivated by the simplicity of interpretation of Type I estimation, which significantly eases the exposition. Given the merely illustrative purpose of the discussion, we consider this simplification to be justified, and refer the reader to [44] for the interpretation of prior distributions in Type II estimation.

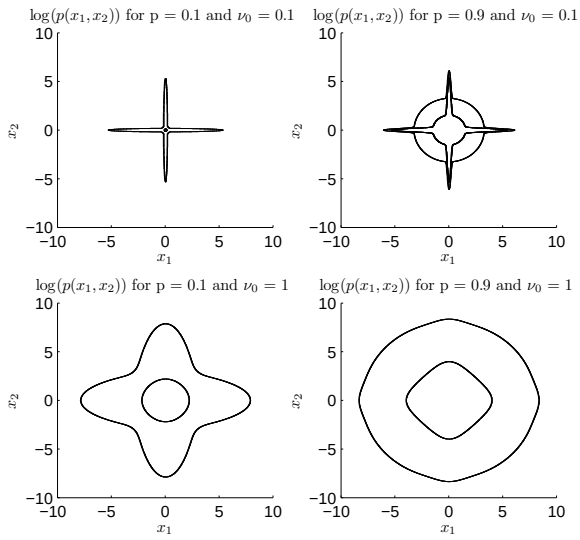


**Fig. 3.3:** Contour plots of the restriction to $\mathbb{R}^2$ of $\log(p(x_1, x_2))$ for two settings of $p$ and $\nu_0$ with $\nu_1 = 3$.

$r_m s_m$ with[5]

$$\begin{aligned} p(r_m) &= \mathrm{CN}(r_m|0, \nu) \\ p(s_m) &= \mathrm{Bern}(s_m|p) \end{aligned} \tag{3.9}$$

where $p \in [0, 1]$ is the rate of the Bernoulli pdf $\mathrm{Bern}(\cdot|p)$. The vector $\mathbf{s}$ collecting all $s_m$ is called the support of the signal: $s_m = 1$ denotes the presence of the signal at position $m$ and $s_m = 0$ denotes its absence, and $\nu$ gives the variance of the non-zero entries of $\mathbf{x}$.

An alternative version of the Bernoulli-Gaussian model is that employed in [48]: $p(\mathbf{x}, \mathbf{s}) = \prod_{m=1}^{M} p(x_m|s_m)p(s_m)$, $p(x_m|s_m) = \mathrm{CN}(x_m|0, \nu_{s_m})$ and $p(s_m) = \mathrm{Bern}(s_m|p)$ with rate $p \in [0, 1]$. Fig. 3.3 shows how different values of $p$ and $\nu_0$ in this model influence the sparsity of $\mathbf{x}$. Decreasing $p$ and $\nu_0$ leads to $p(\mathbf{x})$ being more concentrated along the axes.

With either choice of priors, one can then estimate $\boldsymbol{\theta} \in \{\boldsymbol{\gamma}, \mathbf{s}, \mathbf{r}\}$ based on which estimates of $\mathbf{x}$ are computed. In the following we give an example of sparse signal reconstruction. To derive the algorithm described below, we employ the message-passing formulation of the mean-field approximation introduced in Section 1 of this chapter.

[5]The entries $s_m$ are drawn independently, as are the entries $r_m, m = 1 \dots M$.

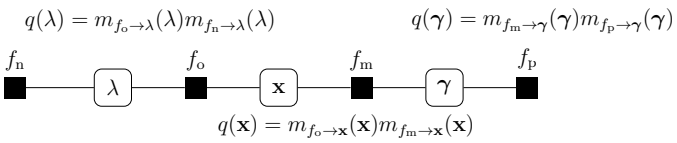


**Fig. 3.4:** Factor graph representation of $p(\mathbf{x}, \lambda, \boldsymbol{\gamma}|\mathbf{y})$ and expression for the beliefs of $\mathbf{x}, \lambda, \boldsymbol{\gamma}$.

**Variational SBI example:** We consider the system (3.7) in which $p(\mathbf{y}|\mathbf{x}, \lambda) = \mathrm{CN}(\mathbf{y}|\boldsymbol{\Phi}\mathbf{x}, \lambda^{-1}\mathbf{I}_N)$ where $\lambda > 0$ is the unknown noise precision. We select the hierarchical prior (3.8) and choose an improper prior for the noise precision [49]: $p(\lambda) \propto 1/\lambda$. We then express the posterior pdf as $p(\mathbf{x}, \lambda, \boldsymbol{\gamma}|\mathbf{y}) \propto f_o(\mathbf{x}, \lambda) f_m(\mathbf{x}, \boldsymbol{\gamma}) f_n(\lambda) f_p(\boldsymbol{\gamma})$ where $f_o(\mathbf{x}, \lambda) = p(\mathbf{y}|\mathbf{x}, \lambda), f_m(\mathbf{x}, \boldsymbol{\gamma}) = p(\mathbf{x}|\boldsymbol{\gamma})$, $f_n(\lambda) = p(\lambda), f_p(\boldsymbol{\gamma}) = p(\boldsymbol{\gamma})$. The posterior pdf has the factor graph representation from Fig. 3.4. Using mean-field update (3.3) for the messages passed from the factor nodes in the factor graph in Fig. 3.4, we compute the beliefs of $\boldsymbol{\gamma}$, $\mathbf{x}$ and $\lambda$ using (3.6): $q(\boldsymbol{\gamma}) \propto \prod_{m=1}^{M} \mathrm{Gam}(\gamma_m|a+1, b+\langle |x_m|^2 \rangle_{\mathbf{x}})$, $q(\lambda) \propto \mathrm{Gam}(\lambda|N, \langle \| \mathbf{y} - \boldsymbol{\Phi}\mathbf{x} \|_2^2 \rangle_{\mathbf{x}})$ and $q(\mathbf{x}) \propto \mathrm{CN}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}})$ where $\boldsymbol{\Sigma}_{\mathbf{x}} = \left( \langle \lambda \rangle_{\lambda} \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Phi} + \langle \mathrm{diag}(\boldsymbol{\gamma}) \rangle_{\boldsymbol{\gamma}} \right)^{-1}$ and $\boldsymbol{\mu}_{\mathbf{x}} = \langle \lambda \rangle_{\lambda} \boldsymbol{\Sigma}_{\mathbf{x}} \boldsymbol{\Phi}^{\mathrm{H}} \mathbf{y}$. [6] Iterating between the updates above is performed until some convergence criterion is fulfilled, then an estimate $\hat{\mathbf{x}} = \boldsymbol{\mu}_{\mathbf{x}}$ is returned. Fig 3.5 shows how different settings for $a$ and $b$ yield solutions with different diversities[7] $d_{\hat{x}}$. We observe that as $a$ increases and/or $b$ decreases, the diversity of $\hat{\mathbf{x}}$ decreases, since the prior $p(\mathbf{x})$ becomes more concentrated along the axes as discussed previously.

## 2.2 Multiple measurement vector (MMV) model

Reconstructing sparse signals in the single measurement vector model (3.7) is an extensively studied problem and many algorithms which exploit the signal sparsity have been developed. However, when $P > 1$ observation vectors sharing the same sparsity structure are available, instead of solving $P$ reconstruction problems, one can benefit from exploiting the "sparsity structure" of the signals of interest and formulate a joint estimation problem [50–52] in the model: $\mathbf{Y} = \boldsymbol{\Phi}\mathbf{X} + \mathbf{W}$ where $\mathbf{Y} \in \mathbb{C}^{N \times P}$ is the measurement matrix consisting of $P$ measurement vectors[8] $\mathbf{Y}_{\cdot,p}, p = 1, ..., P$, while $\mathbf{X} \in \mathbb{C}^{M \times P}$ is the unknown source matrix, each row representing a source vector, and $\mathbf{W}$ is the noise matrix. The source matrix exhibits row sparsity as a result of two important characteristics: (i) the column vectors $\mathbf{X}_{\cdot,p}, p = 1, ..., P$ are sparse and

[6] By $\langle f(\mathbf{x}) \rangle_{\mathbf{x}}$ we denote the expected value of $f(\mathbf{x})$ w.r.t. to the pdf of $\mathbf{x}$.

[7] Following [43], we define the diversity of $\mathbf{x} \in \mathbb{C}^M$ as the cardinality of the set containing all entries of $\mathbf{x}$ above the noise floor: $d_x = |\{x_m : |x_m|^2 > \lambda^{-1}, 1 \leq m \leq M\}|$.

[8] We use $\mathbf{Y}_{\cdot,p}$ and $\mathbf{Y}_{n,\cdot}$ to denote the $p$th column and $n$th row of $\mathbf{Y}$ respectively.

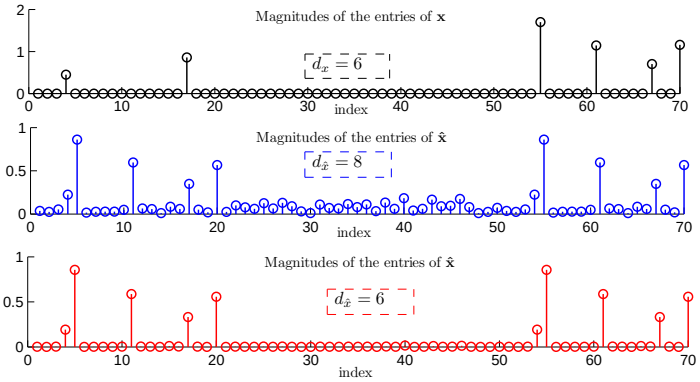


**Fig. 3.5:** Magnitudes of the entries of the sparse signal $\mathbf{x}$ and of the estimated $\hat{\mathbf{x}}$ for two different parameter settings: $a = 0.01, b = 1$ (blue plot) and $a = 1, b = 0$ (red plot). To obtain the observations $\mathbf{y}$ in (3.7), we generate the dictionary $\mathbf{\Phi}$ with entries $\Phi_{n,m} = \exp(-j2\pi nm/M), n \in [0 : N-1], m \in [0 : M-1]$, for $N = 50, M = 70$ and the vectors $\mathbf{w} \sim \mathcal{CN}(0, \lambda^{-1}\mathbf{I}_M)$ with $\lambda = 10^2$ and $\mathbf{x} = \text{diag}(\mathbf{s})\mathbf{r}$, where $p(\mathbf{r}) = \text{CN}(\mathbf{r}|0, \mathbf{I}_M)$ and the entries of $\mathbf{s}$ are drawn independently from a Bernoulli distribution with rate $p = 0.1$.

(ii) they show a common sparsity structure [50] implying that thier non-zeros entries are at the same positions. The probabilistic modeling of the system typically relies on the assumption that the sources $\mathbf{X}_{m,\cdot}, m = 1, ..., M$ are mutually independent and so, a sparsity inducing prior similar to the SMV model is employed (e.g. $p(\mathbf{X}_{m,\cdot}|\gamma_m) = \text{CN}(\mathbf{X}_{m,\cdot}|0, \gamma_m^{-1}\mathbf{B})$ and $p(\gamma_m)$ as in (3.8)). The MMV model can then be recast as a SMV model by applying the transformations $\mathbf{y} = \text{vec}(\mathbf{Y}^{\text{T}})$, $\mathbf{T} = \mathbf{\Phi} \otimes \mathbf{I}_P$, $\mathbf{w} = \text{vec}(\mathbf{W}^{\text{T}})$, and $\mathbf{X} = \text{vec}(\mathbf{X}^{\text{T}})$. [9] Doing so yields $\mathbf{y} = \mathbf{T}\mathbf{x} + \mathbf{w}$. Algorithms applying the SBI methodology can then be used to estimate $\mathbf{x} = [\mathbf{X}_{1,\cdot}, ..., \mathbf{X}_{M,\cdot}]^{\text{T}}$ which consists of $M$ blocks with intra-block correlation determined by the covariance matrix $\mathbf{B}$.

# 3 Wireless communications applications

In this section we give an overview of applications which use the tools introduced in the previous sections.

## 3.1 Iterative receiver design

MF has been successfully applied to the design of receivers for wireless communications, specifically in the design of channel estimators. We note here the works [53–56] which propose estimators of either the impulse [53] or fre-

[9] The operator $\text{vec}(\mathbf{A})$ returns a column vector consisting of all columns of $\mathbf{A}$ stacked on top of one another.

quency [55] response of the channel, or investigating how different belief factorizations impact the accuracy of the estimates [53].

Initially applied to decoding of convolutional [57] or turbo [58] codes, BP gained popularity in recent years. Gaussian approximations of the classical BP messages have been proposed and employed in the design of algorithms for joint detection and channel estimation [59].

A typical application of MF-BP is to the design of iterative algorithms for joint decoding and channel estimation in [60–62], [35, Section IV]. A common feature of these algorithms is that channel response and noise precision estimation is performed by applying MF, while demodulation and decoding is performed using BP. The method allows for a flexible design of the receiver algorithms with various degrees of complexity, i.e. one iteration may consist of updating all or only a specific subset of the unknown variables of the system. Joint iterative receivers that resort to similar inference schemes have been proposed in [63–65].

We find MF-BP particularly suitable for designing receiver algorithms of systems undergoing harsh interference conditions. The method allows for the incorporation of sparsity-inducing priors for the channel weights, enabling therefore the design of sparse channel estimators. By applying MF-BP we are able to approximate the posterior densities of the following variables in the probabilistic system model: the noise precision, the vector of channel weights, the vector of their variances, the modulated data symbols, the coded bits, and the information (uncoded) bits. Consequently, the algorithms embed soft-in soft-out decoding. Yielding soft symbol estimates is a key feature that enables successful interference cancellation because it allows for the data-aided channel estimator to account for the uncertainty in the estimates of the data symbols at any given iteration. In the initial iterations, we expect the symbol decisions to be highly inaccurate, due to the impossibility to cancel the interference before performing a first detection round. Such uncertainty in the symbol decisions is accounted for in the channel estimation phase through the step of re-learning the dictionary matrix (see papers A – C in Part II).

## 3.2 Channel estimation

Since in many propagation environments the channel impulse response is sparse [24], the application of SBI methods to channel estimation (see [22, 24, 66, 67] and the references therein) gained popularity in recent years. To formulate the estimation problem, the typical approach involves as a first step several approximations of the system model which we will discuss next. SBI is then applied to the approximate model.

In the following, we take as a study case the typical pilot-based OFDM channel estimation in which $P$ known complex symbols called pilots, occupying specific time-frequency resources are sent over the wireless channel. By $\mathbf{x} \in$

$\mathbb{C}^P$ we denote the vector collecting all the pilot symbols transmitted during one OFDM symbol and by $\mathbf{f} = [f_{p_1}..f_{p_K}]$ the vector collecting all subcarrier frequencies used to transmit the pilots. We consider the sparse channel with impulse response

$$g(\tau) = \sum_{l=1}^{L} \beta_l \delta(\tau - \tau_l) \tag{3.10}$$

consisting of a small number $L$ of multipath components, with the $l$th component having the gain $\beta_l$ and delay $\tau_l$. For the purpose of this discussion we assume that the channel is invariant during one OFDM symbol.

The pilot-based channel estimator needs to reconstruct the channel frequency response at all subcarrier frequencies based on observing the signals on the $P$ pilot subcarriers collected in the vector $\mathbf{y} \in \mathbb{C}^P$ which is processed to yield a vector $\mathbf{t} = \mathrm{diag}(\mathbf{x})^{-1}\mathbf{y}$ reading

$$\mathbf{t} = \mathbf{h}^{(p)} + \mathbf{w} = \mathbf{A}\boldsymbol{\beta} + \mathbf{w} \tag{3.11}$$

with $A_{k,l} = \exp(-2\pi j f_{p_k} \tau_l), k \in [1:P], l \in [1:L]$, $\mathbf{h}^{(p)}$ is the channel frequency response at pilot positions, $\mathbf{w}$ is the noise vector and $\mathbf{A} \in \mathbb{C}^{P \times L}$ is the dictionary matrix. If the delay vector $\boldsymbol{\tau}$ were known, we could construct the dictionary matrix and compute estimates $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ (using for example the least squares estimator). Then, the estimate of the frequency response $h_i$ at a subcarrier frequency $f_i$ could be computed as $\hat{h}_i = \sum_{l=1}^{L} \hat{\beta}_l \exp(-2\pi j f_i \tau_l)$.

However, since $\boldsymbol{\tau}$ is typically unknown, an approximation $\tilde{\mathbf{h}}$ of the true $\mathbf{h}$ is derived instead. This is usually accomplished by imposing a dense delay grid $\boldsymbol{\xi} = \{\xi_l = (l-1)\Delta\tau; l \in [1:K]\}$ with resolution $\Delta\tau$, length $K$, and a vector $\boldsymbol{\alpha} \in \mathbb{C}^K$ so that the approximate channel frequency response at $i$th subcarrier is $\tilde{h}_i = \sum_{l=1}^{K} \alpha_l \exp(-2\pi j f_i \xi_l)$. The response $\tilde{\mathbf{h}}$ is a good approximation of $\mathbf{h}$ if a sufficiently fine resolution $\Delta\tau$ and a length $K \gg L$ are selected. Then, due to the channel sparsity, we expect that the approximation holds for a vector $\boldsymbol{\alpha}$ having many zero entries. Consequently, we can now write the standard model for compressed sensing:

$$\mathbf{t} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{w} \tag{3.12}$$

with $D_{k,l} = \exp(-2\pi j f_{p_k} \xi_l), k \in [1:P], l \in [1:K]$. We can then use SBI to compute sparse estimates of the vector $\boldsymbol{\alpha}$ and subsequently, of the vector $\tilde{\mathbf{h}}$.

Note that in a practical communication system the matrix $\mathbf{A}$ in (3.11) is no longer a Fourier dictionary since the channel frequency response experienced by the receiver contains effects induced by the transceiver structure e.g. transmit and receive filters. While the receive filters responses may be known and can be accounted for when designing the dictionary, the same thing cannot be said about the transmit filters responses; hence, some additional assumptions need to be made to facilitate the design of an approximate dictionary $\mathbf{D}$. We

investigate the effect of the mismatch between dictionaries $\mathbf{A}$ and $\mathbf{D}$ in paper D of Part II.

Once the expression of the approximate signal (3.12) is derived, the probabilistic model can be next cast. SBI channel estimators using either of the two hierarchical models introduced earlier (Gamma-Gaussian [53, 66] or Bernoulli-Gaussian [68, 69]) can be designed to estimate $\boldsymbol{\alpha}$. Due to their Bayesian formulation, these estimators are easily incorporated in iterative receiver schemes implementing approximate Bayesian inference [68, 70] that perform joint channel estimation and data detection. The estimation of the MIMO subchannels with sparse common support is solved in [27] by transforming the MMV model into an SMV model and applying the algorithm in [38], while the estimation of sparse time-varying channels from correlated observation vectors is tackled in [71].

# Chapter 4

# Thesis Contributions and Outlook

In this chapter we detail the contributions of the thesis and provide an outlook of this work. This chapter is structured as follows: Section 1 briefly introduces the main contribution of each paper (a full version of which can be found in Part II) and Section 2 sketches the patents – filed or granted – generated within this project.

## 1 Research articles

The research articles presented in the following propose various receiver designs for the harsh interference conditions described in Chapter 2 and channel estimation techniques, utilizing the tools described in Chapter 3.

### Paper A: Message-passing receiver for OFDM systems over highly delay-dispersive channels

In this paper we propose an algorithm for receivers operating in insufficient CP conditions. The algorithm is derived by applying MF-BP to obtain a joint design of the following tasks: data-aided estimation of the sparse channel impulse response and the noise variance, reconstruction and cancellation of intersymbol and intercarrier interference, and decoding over a block of symbols. The variational Bayesian framework that we employ allows us to easily incorporate sparsity constraints on the prior of the channel weights to obtain sparse estimates of them.

Numerical evaluations show that the receiver implementing our algorithm exhibits good decoding performance compared with the reference receivers, jus-

tifying therefore the approximations and modeling choices that we resorted to in our design. The good performance is also an indicator that intercarrier and intersymbol interference can be efficiently mitigated. Therefore a reduction or even a removal of the cyclic prefix at the benefit of increased spectral efficiency might be a sensible 5G strategy. Finally, the algorithm is versatile since it pertains to various sizes of the processed block, schedules for updating the different quantities of interest, channel estimation strategies. This virtue enables the tunning of the algorithm for either performance boosting or complexity reduction.

## Paper B: OFDM receiver for fast time-varying channels using block-sparse Bayesian learning

In response to the problem risen in Section 2 of Chapter 2, we design an algorithm for receivers operating in high mobility scenarios where the orthogonality of the subcarriers cannot be preserved. To formulate the channel estimation problem, we employ a basis expansion model to characterize the time-variant behavior of each multipath component. This model yields a fixed small number of unknown coefficients associated with each channel weight. Then, to approximate the time-variation of each component, and hence the intercarrier interference, it suffices to estimate these coefficients. However, under the assumption of a sparse channel, many components are negligible, and so are all associated coefficients. The vector collecting all these coefficients will therefore exhibit many zero entries occurring in blocks, i.e. will be block-sparse. This particular pattern allows us to formulate the channel estimation problem as one of reconstructing block-sparse signals. As in Paper A, we employ MF-BP to formulate a receiver algorithm performing block-sparse, data-aided channel estimation on the above approximate model, along with intercarrier interference cancellation and data decoding.

The numerical study shows that the receiver implementing our algorithm successfully outperforms the reference receivers. Its very good performance validates the model described above. This is also an indicator that very-high mobility 5G users can operate well, as long as they are equipped with algorithms tailored to account for the large Doppler spread of the fast time-varying channels.

## Paper C: Interference-aware OFDM receiver for channels with sparse common supports

In this paper we investigate different OFDM receiver implementations for operation in co-channel interference conditions. We design a MF-BP algorithm that performs joint channel estimation, interference cancellation and decoding. To that end we rely on two assumptions: that the MIMO-subchannel responses are

sparse and those associated to a same transmit antenna have a common support. The latter assumption stems from the fact that the number of multipath components and respective delays of the channel responses corresponding to a given transmitter and all receive antennas are expected to be equal, since these antennas are collocated. The channel estimation problem, of both desired and interfering channels, is formulated as one of reconstructing sparse signals from multiple measurement vectors that share the same unknown sparse support. It is solved using sparse Bayesian inference.

In this paper we explore through simulations how the proposed design and the information about the interfering signal impact the overall receiver performance in different signal-to-interference power regimes. Our results show that a receiver which knows the modulation and code of the interferer has excellent interference cancellation capabilities, and hence granting access to this information can be a useful 5G feature. An encouraging finding is that even when receiver makes mismatched assumptions about these parameters, it still outperforms interference-unaware receivers, and has good performance in medium signal-to-interference regime. To conclude, enabling the receiver to estimate the interfering channels has clear benefits. However, there are several aspects that need to be taken into account such as reference signals design, inter-cell synchronization, etc.

## Paper D: Sparse channel estimation including the impact of the transceiver filters with application to OFDM

Traditionally, the dictionary matrix used in the estimation of sparse wireless channels is obtained from the discrete Fourier transform, based on the assumption that the channel frequency response can be approximated as a linear combination of a small number of multipath components. In practical communication systems, however, the channel response observed at the receiver includes additional effects to those induced by the propagation channel. This composite channel embodies, in particular, the impact of the transmit and receive filters. Hence, the assumption of the channel frequency response being sparse in the classical Fourier dictionary becomes merely a (coarse) approximation.

In this work, we derive a signal model, and subsequently a novel dictionary matrix, for sparse estimation that accounts for the impact of transceiver filters. Employing the proposed dictionary proves beneficial to the performance of the receiver and it is particularly suited for the operation over sparse channels. Based on the results of this study we conclude that using a dictionary including the responses of the transceiver filters results in more accurate estimates of the channel frequency response, and that compressed channel estimation is a promising technique which 5G systems could benefit from.

# 2 Industrial innovation

Herein we list and sketch the patents that have been generated based on the ideas presented in the papers reported in Section 1.

## Patent E: Methods and devices for channel estimation and OFDM receiver

This patent has been filed in relation to paper A. The patent describes the architecture of an OFDM receiver able to operate over channels with responses characterized by a maximum excess delay larger than the cyclic prefix advertised by the serving base station. The patent specifies how the channel estimation and decoding should be performed in order to cope with the resulting intercarrier and intersymbol interference.

## Patent F: Methods and devices for channel estimation for mobile systems of insufficient cyclic prefix length

Also filed in relation to paper A, this patent proposes a pilot-based channel estimator for an OFDM receiver operating in insufficient CP conditions. The estimator uses an approximate signal model to reconstruct and cancel the interference and estimate the gains and the delays of the multipath components. The design of the algorithm relies on the application of methods from sparse Bayesian inference.

## Patent G: Devices and methods for processing one or more received radio signals

This patent describes the architecture of an OFDM receiver able to operate in highly Doppler-dispersive channels. The architecture uses the estimation model developed in paper B and relies on interconnecting the channel estimator, the equalizer and the detector to enable the reconstruction and cancellation of interference both when estimating the channel responses and when decoding the data.

## Patent H: Channel estimation technique

Developing the idea presented in paper D, this granted patent describes a pilot-based channel estimator for an OFDM receiver that computes sparse estimates of the channel impulse response. The design of the channel estimation block incorporates the responses of the transceiver filters, making it thus device-dependent.

# 3 Conclusions and outlook

In this thesis we propose algorithms for enabling the operation of the OFDM receiver in severe interference conditions. Based on the technological trends that are driving the definition of upcoming 5G systems, we expect that such severe operating conditions will become increasingly relevant in such systems.

To reiterate, we investigate interference scenarios arising due to a) the channel maximum excess delay exceeding the duration of the cyclic prefix, b) the high mobility of the user equipment and c) the usage of the same time-frequency resources by multiple users. For each of these situations, we derive joint channel estimation, equalization and decoding schemes for interference-aware receivers. The design of these schemes relies on a probabilistic modeling of the system under the conditions a), b) or c) which exploits the sparsity of the wireless propagation channel and on the application of MF-BP, the latter method enabling a unified design of all receiver's tasks.

Exploiting the sparsity of the channel to estimate its frequency response is accomplished in a two-step method: first, to avoid the estimation of the multipath components delays, we impose the delays to be constrained on a fixed grid with sufficiently fine resolution so that the channel frequency response is well approximated and second, we impose hierarchical models over the channel weights which favors sparse solutions. In this context, we investigate sparse Bayesian estimators using two such models,: a Gamma-Gaussian prior and a Bernoulli-Gaussian prior. Applying MF-BP enables the incorporation of the channel estimator in a joint Bayesian receiver design. The additional benefits of using this framework are 1) that the estimation uncertainties (in the estimated channel, desired bit stream) are accounted for when reconstructing the interference and 2) the flexibility of the iterative algorithm whose parameters can be tuned to allow for trading complexity with performance. Even though the improved performance of the proposed receiver algorithms comes at the price of a higher complexity, given the steady increase of computation capability of the modem, we expect that practical implementations of these algorithms will become feasible in the future.

In order to understand the effects of a), b) and c) on the operation of the wireless receiver, we treat each problem independently and formulate signal models and algorithms designed for the conditions imposed in each scenario. Although it simplifies the channel estimation, the gridded-delay approach yields a mismatched model of the signal used in the algorithm design relative to the true model. Including the delay estimation in the algorithm design is a feature that could be beneficial to the overall receiver performance, at the expense of increasing complexity. Nevertheless, it constitutes a research question worth investigating. Formulating a prior of the channel weights which mimics the channel property to exhibit clustering of multipath components in delay domain is another research avenue to be considered when designing sparse channel

estimators.

In this work we choose OFDM as a base technology for our study. Two main reasons have justified our choice: it is a well understood technology, with a worldwide adoption that is heading the competition for the 5G waveform, and all other 5G contenders ( see [72] for details) are "incremental departures from OFDM" [3]. However, a further study could consists of the development of algorithms for the joint treatment of a), b) and c) and tailored to a multi-carrier system which offers better spectral efficiency, e.g. by relaxing the strict subcarrier orthogonality.

We emphasize the importance of studying receiver designs for harsh interference conditions in 4G-like systems. The conclusions of each study do not only offer valuable insights into the design of receivers for 5G systems but also help formulate guidelines for the design of these systems, the latter having direct consequences on the performance and complexity of 5G receivers.

Lastly, this project has been conducted as an industry-academia cooperation. The outcomes of the studies performed throughout the project provide insights and guidelines meant to help the modem evolution towards enhanced baseband algorithms that can ensure the operation of the user equipment in extreme conditions. The operation of such equipment however is conditioned on the overall design of the communication system, a matter addressed by standardization bodies. Therefore, coordinated efforts from both sides (standardization community and chip manufacturers) are key to enabling designs of algorithms like those proposed in this work.

# References

[1] "Ten key rules of 5G deployment," Nokia Networks White Paper.

[2] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.

[3] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[4] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[5] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[6] R. Ghaffar and R. Knopp, "Interference-aware receiver structure for Multi-User MIMO and LTE," *"EURASIP Journal on Wireless Communications and Networking", Volume 2011: 40*, 05 2011. [Online]. Available: http://www.eurecom.fr/publication/3394

[7] N. DOCOMO INC, "DOCOMO 5G White paper -5G radio access: Requirements, concept and technologies," 2014.

[8] X. Jun, "Progress on 5G radio link enhancement," ZTE Technologies, Tech. Rep., June 2015.

[9] S. Bittner, E. Zimmermann, and G. Fettweis, "Low complexity soft interference cancellation for MIMO-systems," in *IEEE VTC*, vol. 4, May 2006, pp. 1993–1997.

[10] Z. Chen, C. Yongyu, and D. Yang, "Low-complexity turbo equalization for MIMO-OFDM system without cyclic prefix," in *IEEE PIMRC*, Sept 2009, pp. 310–314.

[11] R. Nissel and M. Rupp, "Doubly-selective mmse channel estimation and ICI mitigation for OFDM systems," in *IEEE ICC*, June 2015, pp. 4692–4697.

[12] P. Schniter, "Low-complexity equalization of OFDM in doubly selective channels," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1002–1011, April 2004.

[13] L. Rugini, P. Banelli, and G. Leus, "Simple equalization of time-varying channels for OFDM," *IEEE Commun. Lett.*, vol. 9, no. 7, pp. 619–621, July 2005.

[14] A. Molisch, M. Toeltsch, and S. Vermani, "Iterative methods for cancellation of intercarrier interference in OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 2158–2167, July 2007.

[15] K. Fang, L. Rugini, and G. Leus, "Low-complexity block turbo equalization for OFDM systems in time-varying channels," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5555–5566, Nov 2008.

[16] A. Gorokhov and J.-P. Linnartz, "Robust OFDM receivers for dispersive time-varying channels: equalization and channel acquisition," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 572–583, April 2004.

[17] E. Panayirci, H. Senol, and H. Poor, "Joint channel estimation, equalization, and data detection for OFDM systems in the presence of very high mobility," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4225–4238, Aug 2010.

[18] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Trans. Commun.*, vol. 52, no. 2, pp. 82–89, February 2014.

[19] M.-A. Badiu, M. Guillaud, and B. Fleury, "Interference alignment using variational mean field annealing," in *WiOpt*, May 2014, pp. 585–590.

[20] S. Wagner and F. Kaltenberger, "Interference-aware receiver design for MU-MIMO in LTE: Real-time performance measurements," *Intel Technology Journal, Volume 18, No 1, 2014*, 01 2014. [Online]. Available: http://www.eurecom.fr/publication/4213

[21] L. Yang and W. Zhang, *Interference Coordination for 5G Cellular Networks*, ser. SpringerBriefs in Electrical and Computer Engineering. Springer International Publishing, 2015. [Online]. Available: https://books.google.de/books?id=dwDOCgAAQBAJ

[22] C. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Magazine*, vol. 48, no. 11, pp. 164–174, November 2010.

[23] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, "Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255–271, April 2010.

[24] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, June 2010.

[25] J. J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in ofdm systems," in *IEEE Vehicular Technology Conference*, vol. 2, Jul 1995, pp. 815–819 vol.2.

[26] G. Gui, L. Xu, and L. Shan, "Block Bayesian sparse learning algorithms with application to estimating channels in OFDM systems," in *WPMC*, Sept 2014, pp. 238–242.

[27] R. Prasad and C. Murthy, "Joint channel estimation and data detection in MIMO-OFDM systems using sparse Bayesian learning," in *NCC*, Feb 2014, pp. 1–6.

[28] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, Jul 2000.

[29] O. Edfors, M. Sandell, J.-J. van de Beek, S. Wilson, and P. Borjesson, "OFDM channel estimation by singular value decomposition," in *IEEE VTC*, vol. 2, Apr 1996, pp. 923–927 vol.2.

[30] N. Pedersen, C. Navarro Manchón, D. Shutin, and B. Fleury, "Application of Bayesian hierarchical prior modeling to sparse channel estimation," in *IEEE ICC*, June 2012, pp. 3487–3492.

[31] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and trends in Machine Learning*, vol. 1, no. 1–2, June 2008.

[32] J. S. Yedidia, W. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, July 2005.

[33] E. P. Xing, M. I. Jordan, and S. Russell, "A generalized mean field algorithm for variational inference in exponential families," in *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, ser. UAI'03. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003, pp. 583–591. [Online]. Available: http://dl.acm.org/citation.cfm?id=2100584.2100655

[34] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., 1988.

[35] E. Riegler, G. Kirkelund, C. Navarro, M. Badiu, and B. Fleury, "Merging belief propagation and the mean field approximation: A Free Energy Approach," *IEEE Trans. Inf. Theory*, vol. 59, pp. 588–602, 2013.

[36] J. Dauwels, "On variational message passing on factor graphs," in *IEEE ISIT*, June 2007, pp. 2546–2550.

[37] T. Heskes, "Stable fixed points of loopy belief propagation are local minima of the bethe free energy," in *Advances in Neural Information Processing Systems 15*, S. Thrun and K. Obermayer, Eds. Cambridge, MA: MIT Press, 2002, pp. 343–350. [Online]. Available: http://books.nips.cc/papers/files/nips15/LT04.pdf

[38] M. E. Tipping and A. Smola, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, 2001.

[39] M. W. Seeger, "Bayesian inference and optimal design for the sparse linear model," *J. Mach. Learn. Res.*, vol. 9, pp. 759–813, Jun. 2008. [Online]. Available: http://dl.acm.org/citation.cfm?id=1390681.1390707

[40] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[41] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[42] M. A. T. Figueiredo, "Adaptive sparseness using Jeffreys prior," in *Advances in Neural Information Processing Systems 14*. MIT Press, 2001, pp. 697–704.

[43] D. P. Wipf, "Bayesian methods for finding sparse representations," Tech. Rep., 2006.

[44] D. P. Wipf, B. D. Rao, and S. Nagarajan, "Latent variable bayesian models for promoting sparsity," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6236–6255, Sept 2011.

[45] J. Palmer, B. D. Rao, and D. P. Wipf, "Perspectives on sparse Bayesian learning," in *Advances in Neural Information Processing Systems*, S. Thrun, L. Saul, and B. Schölkopf, Eds. MIT Press, 2004, pp. 249–256. [Online]. Available: http://papers.nips.cc/paper/2393-perspectives-on-sparse-bayesian-learning.pdf

[46] N. L. Pedersen, C. N. Manchón, M.-A. Badiu, D. Shutin, and B. H. Fleury, "Sparse estimation using Bayesian hierarchical prior modeling for real and complex linear models," *Signal Processing*, vol. 115, pp. 94 – 109, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168415001140

[47] J. J. Kormylo and J. Mendel, "Maximum likelihood detection and estimation of bernoulli - gaussian processes," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, pp. 482–488, May 1982.

[48] P. Schniter, L. Potter, and J. Ziniel, "Fast Bayesian matching pursuit," in *Information Theory and Applications Workshop*, Jan 2008, pp. 326–333.

[49] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 92–106, Jan 2009.

[50] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, July 2005.

[51] Z. Zhang and B. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 912–926, Sept 2011.

[52] D. Wipf and B. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, July 2007.

[53] N. L. Pedersen, C. Navarro Manchón, and B. Fleury, "Low complexity sparse Bayesian learning for channel estimation using generalized mean field," in *European Wireless*, May 2014, pp. 1–6.

[54] M. Badiu, C. Navarro Manchón, V. Bota, and B. Fleury, "Distributed iterative processing for interference channels with receiver cooperation," in *ISTC*, Aug 2012, pp. 140–144.

[55] C. Navarro Manchón, B. Fleury, G. Kirkelund, P. Mogensen, L. Deneire, T. Sorensen, and C. Rom, "Channel estimation based on divergence minimization for OFDM systems with co-channel interference," in *IEEE ICC*, June 2009, pp. 1–6.

[56] B. Hu, I. Land, L. Rasmussen, R. Piton, and B. Fleury, "A divergence minimization approach to joint multiuser decoding for coded cdma," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 432–445, April 2008.

[57] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, pp. 498–519, 2001.

[58] R. McEliece, D. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of Pearl's belief propagation algorithm," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 2, pp. 140–152, Feb 1998.

[59] C. Novak, G. Matz, and F. Hlawatsch, "Factor graph based design of an OFDM-IDMA receiver performing joint data detection, channel estimation, and channel length selection," in *IEEE ICASSP*, April 2009, pp. 2561–2564.

[60] C. Navarro Manchón, G. E. Kirkelund, E. Riegler, L. P. B. Christensen, and B. H. Fleury, "Receiver architectures for MIMO-OFDM based on a combined VMP-SP algorithm," *ArXiv 1111.5848*, 2011.

[61] M.-A. Badiu, C. Navarro Manchón, and B. Fleury, "Message-passing receiver architecture with reduced-complexity channel estimation," *IEEE Commun. Lett.*, vol. 17, pp. 1404–1407, 2013.

[62] C. Zhang, C. Navarro Manchón, Z. Wang, and B. Fleury, "Message-passing receivers for single carrier systems with frequency-domain equalization," *IEEE Signal Process. Lett.*, vol. 22, no. 4, pp. 404–407, April 2015.

[63] M. Tuchler and A. Singer, "Turbo Equalization: An Overview," *IEEE Trans. Inf. Theory*, vol. 57, pp. 920–952, 2011.

[64] Q. Guo, L. Ping, and D. Huang, "A low-complexity iterative channel estimation and detection technique for doubly selective channels," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 4340–4349, 2009.

[65] Y. Zhu, D. Guo, and M. Honig, "A message-passing approach for joint channel estimation, interference mitigation, and decoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 6008–6018, December 2009.

[66] R. Prasad, C. Murthy, and B. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3591–3603, July 2014.

[67] S.-J. Hwang and P. Schniter, "Efficient multicarrier communication for highly spread underwater acoustic channels," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 9, 2008.

[68] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," in *IEEE SPAWC*, June 2011, pp. 246–250.

[69] J. Ziniel and P. Schniter, "Dynamic compressive sensing of time-varying signals via approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5270–5284, Nov 2013.

[70] P. Schniter, "Joint estimation and decoding for sparse channels via relaxed belief propagation," in *Asilomar*, Nov 2010, pp. 1055–1059.

[71] J. W. Choi and B. Shim, "Statistical recovery of simultaneously sparse time-varying signals from multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 6136–6148, Nov 2015.

[72] A. Sahin, I. Guvenc, and H. Arslan, "A survey on multicarrier communications: Prototype filters, lattice structures, and implementation aspects," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1312–1338, 2014.

# Part II

# Papers

# Paper A

## Message-passing Receiver for OFDM Systems over Highly Delay-Dispersive Channels

Oana-Elena Barbu, Carles Navarro Manchón, Christian Rom, Bernard H. Fleury

The paper has been submitted to
*IEEE Transactions on Wireless Communications*
March, 2016

# Abstract

*Propagation channels with maximum excess delay exceeding the duration of the cyclic prefix of wireless OFDM-based systems – like 4G and potentially its successors – cause intersymbol and intercarrier interference which, unless accounted for, drastically impair the receiver performance.*

*Using tools from variational Bayesian inference and sparse signal reconstruction, we derive an iterative algorithm that performs estimation of the channel impulse response and the noise variance, interference reconstruction and cancellation and data decoding over a block of symbols.*

*Simulation results show that the receiver employing our algorithm outperforms receivers applying traditional interference cancellation and pilot-based schemes, and it approaches the performance of an ideal receiver with perfect channel state information and perfect interference cancellation capabilities.*

*We highlight the relevance of our algorithm and the related study in the context of both 4G/LTE and 5G. By enabling the 4G receiver experiencing the above harsh conditions to locally cancel the interference, our design circumvents the spectral efficiency loss incurred by extending the duration of the cyclic prefix – a handy yet dissipative solution. Furthermore, it sets the premises for the development of advanced receivers for future multicarrier systems like 5G in which the strict subcarrier orthogonality will most likely be relaxed.*


# 1 Introduction

## 1.1 Motivation

### Problem statement

Multipath propagation in the wireless channel causes multiple copies of the transmitted signals to arrive at the receiver with distinct weights, delays, directions and with distinct Doppler frequencies in time-varying conditions. If not appropriately dealt with, this phenomenon causes intersymbol interference (ISI) in the received signal. Current OFDM systems preempt ISI by prepending to each OFDM symbol a cyclic prefix (CP) whose duration is selected (and fixed) larger than the typical[1] maximum excess delay (MED) of the channels in which the systems are designed to operate. When this condition is fulfilled, the observed signal at each subcarrier consists of the symbol transmitted on that subcarrier multiplied by the channel gain at the subcarrier frequency plus additive white Gaussian noise (AWGN). This is the traditional signal model employed in the design of OFDM receivers. Since the CP insertion comes

[1]By typical MED, we mean the MED specified in standardized vehicular and pedestrian channels from [1] that are widely used in the design and testing of wireless modems.

at the cost of reduced spectral efficiency, it is highly inadvisable to extend its duration to cope with "worst-case" propagation conditions [2]. Therefore, when this situation occurs, the receiver is forced to operate over a channel with MED larger than the CP duration, i.e. in the so-called insufficient CP condition. In this case, the receiver experiences a signal impaired by both ISI and inter-carrier interference (ICI) [3], generically called self-interference, causing an overall degradation of the system performance [2, 4, 5]. Specifically, the observed signal at each subcarrier contains now the noisy contribution of the signals transmitted on all subcarriers during the current and the previous signaling intervals. Receivers designed using the traditional (interference free) signal model will therefore erroneously estimate the channel, and then use this erroneous estimate to poorly equalize and decode the received signals.

**Insufficient CP conditions in 4G/LTE**

Propagation conditions in which the channel MED exceeds the normal CP duration of $4.69\mu s$ [6] occur infrequently but often enough to deserve attention and be accounted for in 4G/LTE. Such conditions have been documented in the literature on several occasions [7–9]. For example, the authors of [7] report channel impulse responses (CIR) with MED $> 15\mu s$ measured in Manchester city center. Similarly, measurement campaigns carried out in Helsinki [9] revealed PDPs with multipath components arriving at $\sim 6.6\mu s$ having the strongest power (see [9, Fig. 5]), leading to a MED of $\sim 7\mu s$. The propagation conditions enumerated above have been classified by the wireless propagation community as *bad urban* (BU) or *hilly terrain* (HT) environments and standard channel models that mimic those conditions have been developed and adopted in the 3GPP standards to test wireless communication systems [8, 10]. Specifically, two propagation models, among others, are of interest for our work, i.e. the typical BU [8, Section 2.4.2.3] and the typical HT [8, Section 2.4.2.4] channel models. The PDPs of both models exhibit two clusters of multipath components leading to a MED exceeding $10\mu s$. The first cluster is the result of propagation via scatterers located between and around the serving base station and the mobile terminal, while the second cluster of non-negligible power is caused by reflections and scattering from tall and distant scatterers.

**System adaptation in 4G/LTE**

A potential solution to cope with such harsh conditions is to allow for an adaptive setting of the system parameters. The authors of [11] investigate several parameters, including the CP duration, that can be adjusted to minimize the performance degradation. However, LTE systems implement a multiple access scheme in which users are multiplexed in time and frequency by transmitting over only a subset of the available subcarriers and OFDM symbols in a given

frame. Hence the base station is forced to select and broadcast a single CP, which cannot be optimal with respect to all the different channel conditions experienced by all served users. Adapting the CP in a conservative manner, i.e. setting a CP large enough to accommodate the users experiencing the longest excess delay, would only result in an unnecessary reduction of spectral efficiency of all users. Instead, to avoid that communication completely fails, the systems possess other kinds of link adaptation mechanisms e.g. lowering the modulation and/or coding rate of the user experiencing insufficient CP conditions, using more retransmissions per packet etc. This approach however discriminates users served by a same base station since those undergoing insufficient CP conditions will experience a lowered quality of service compared to those operating in nominal conditions.

**Outlook towards 5G**

With the advent of 5G systems, the trend observed in the design of waveforms of many candidate OFDM-based systems is towards a removal of the CP or at least a shortening of its duration [12], making the resulting self-interference a more stringent and frequent problem. To cope with it, these systems will instead rely on advanced signal processing algorithms implemented in 5G receivers equipped with sufficient computational power. Such receivers are generically classified as advanced receivers or smart devices [13–15]. They are expected to cancel the self-interference through more complex equalization and detection [16].

Bearing in mind the above considerations, we believe that an algorithm tailor-made for receivers operating in insufficient CP conditions represents a viable option for OFDM-like wireless communication systems. On the one hand, such an algorithm is an enabler towards providing users undergoing worst-case propagation conditions the same quality of service as those experiencing nominal propagation conditions. On the other hand, the design of such an algorithm and the insights that it provides constitute valuable input to the development of spectrally efficient 5G systems. Hence, the relevance of the problem and of the solution that we present in this paper are to be seen within the context of both 4G/LTE and 5G.

## 1.2 Previous work on OFDM receivers for insufficient CP conditions

Most receiver algorithms proposed for OFDM systems operating in insufficient CP conditions consist of an initial ISI cancellation step, followed by iterative ICI cancellation. Among these schemes we mention the operator-perturbation technique (OPT) and parallel/serial interference cancellation (P/SIC) [2]. Iterative methods that remove ISI and reconstruct the OFDM symbols cyclicity have been proposed in [17] and further developed in [18–20]. Other methods use tap-selection algorithms in decision-feedback equalizers (DFE). These al-

gorithms increase the span of the equalizer in sparse channels with large delays spreads [21, 22]. The aforementioned interference cancellation algorithms lead to excellent receiver performance under the assumption of perfect channel state information (CSI) at the receiver. However, little attention has been paid to the problem of how to acquire accurate CSI under insufficient CP conditions. Most works [23, 24] assume perfect CSI to be already available at the receiver. Others [4, 25] propose the use of coarse estimates obtained from pilot-based channel estimators, derived under the assumption that the CIR is well approximated by a truncated baseband representation.

To summarize, current OFDM receivers designed to operate in insufficient CP conditions focus almost entirely on canceling the resulting interference at equalization. They either discard the channel estimation problem or treat it as an independent receiver task, i.e. by assuming that the channel estimates obtained by neglecting the self-interference are sufficiently accurate. However, in a typical OFDM system, like LTE, pilot symbols are multiplexed with data symbols [6]. In insufficient CP conditions, this implies that a) pilots are also corrupted by self-interference, and hence, a pilot-based channel estimation method will perform poorly unless the interference is reconstructed and canceled, and b) the low pilot density does not allow for accurate estimation of the highly frequency-selective channel.

In view of the above, we advocate that an OFDM receiver for insufficient CP conditions should be designed such that all its building blocks are interference aware: ISI and ICI should be accounted both when retrieving the channel weights and when performing data detection.

Receivers that treat jointly the channel estimation and detection tasks have been earlier proposed to cancel the ICI arising from Doppler-dispersion of the channel [26] or impulsive noise [27]. However, these works and our work treat significantly different problems due to the specificities of the considered applications.

## 1.3 Proposed approach and theoretical background

In this paper we propose an iterative algorithm for OFDM receivers operating in insufficient CP conditions. The algorithm performs data-aided channel estimation, ISI and ICI cancellation, noise variance estimation, equalization and decoding by applying tools from variational Bayesian inference (VBI). The channel estimation block of the proposed algorithm is designed under the assumption that the CIR is sparse, i.e. it is composed of a few non-zero multipath components, and hence can be implemented using sparse Bayesian inference (SBI).

SBI is a Bayesian formulation of compressed sensing which encompasses strategies for reconstructing sparse signals by formulating a probabilistic framework of the system and imposing a prior distribution on the signals of interest

which favors sparse solutions [28]. Precisely because of its Bayesian formulation, SBI is particularly well suited for embedding in a Bayesian inference framework. We include it in our probabilistic model by defining sparsity-inducing prior distributions for the weights of the CIR. Similar approaches have been proposed earlier to estimate sparse CIR, but in traditional (interference free) signal models [29] [30]. In fact, the authors of [30] use the expectation-maximization framework to also formulate an iterative algorithm that performs joint channel estimation and symbol detection, however without including decoding as part of the joint structure. To embed soft-in soft-out decoding in the joint design of our interference-aware receiver algorithm, we resort to a particular VBI technique as discussed next.

VBI techniques comprise methods for obtaining approximations of the posterior probability densities – called beliefs – of unobserved variables in a probabilistic model. The use of this type of inference techniques for the design of wireless receivers has become widespread in recent years [31–36]. In this work we use the hybrid framework proposed in [37] which combines the popular belief propagation (BP) [38] and mean-field approximation (MF) [39] – to which we will generically refer to as MF-BP.

## 1.4 Contribution

The main contribution of this article is the derivation of a joint receiver for OFDM systems in insufficient CP conditions. The receiver iteratively performs data-aided, interference-aware channel and noise variance estimation, self-interference reconstruction and cancellation, channel equalization and data decoding. By using MF-BP on the probabilistic model of the communication system, we obtain a unified design of the above tasks. To the authors' knowledge, such an integrated design has not been proposed for OFDM receiver in insufficient CP conditions before. In the following, we detail each of the design traits that enable our solution:

- We develop a mathematical model of the signal received in insufficient CP conditions which allows us to express the resulting self-interference using the complex weights and delays of the CIR and the current and previous transmitted signals. We formulate two (equivalent) representations of the model, one suitable for the design of the detector, and one tailored to the estimation of the channel weights. The latter representation linearly maps the CIR weights to the received signal through a dictionary matrix which includes ISI and ICI. This alternative expression of the received signal is the foundation for a compressed-sensing-based formulation of the channel estimator, which accounts for ISI and ICI.

- To design the channel estimator and detector we derive the corresponding approximate signal models which allow us to circumvent the estimation of

the delays (inherently needed to reconstruct ISI and ICI). Specifically, we assume that the delays of the multipath components lie on a grid with certain resolution. This enables us to model and reconstruct the self-interference without the explicitly estimating the multipath components' delays. By combining the delay-discretization approximation with the assumption that the CIR is sparse, we apply the SBI framework and select a probabilistic model characterizing the system which allows us to obtain tractable solutions.

- To obtain accurate weight estimates, the proposed algorithm includes a dictionary matrix learning step: i.e. after the reconstruction of ISI and ICI, the dictionary matrix is re-computed. This is an essential step which enables the channel estimation block to account for the self-interference when refining the estimates of the channel weights.

- The algorithm is designed within a framework in which a global objective function is optimized. As a result, we obtain a coherent iterative structure where the different constituent blocks operate and exchange information between them following a unified principle.

We evaluate the performance of the proposed algorithm and other state-of-art algorithms using Monte Carlo simulations. Our results reveal that the state-of-art algorithms, which are designed under the assumption of perfect CSI at the receiver, are sensitive to inaccuracies in the channel estimates and experience considerable BER performance degradation. On the contrary, the proposed receiver benefits from the joint design of channel estimator and equalizer, performing significantly closely to ideal receivers having perfect CSI.

The remainder of this paper is organized as follows: in Section 2.1 we develop a model for the OFDM signal received when operating in insufficient CP conditions. In Section 2.2 we derive an approximate signal and probabilistic models tailored to the application of SBI. In Section 3.1 we describe the theoretical foundations of the iterative receiver algorithm described in Section 3.2 – 3.4. In Section 4 we carry out several performance studies in which we compare our receiver algorithm with other schemes. We present our conclusions in Section 5.

*Notation:* By $[1:P]$ we denote the set $\{p \in \mathbb{N} | 1 \leq p \leq P\}$. The $(i,j)$ element of matrix $\mathbf{A}$ is denoted as $A[i,j]$ and $\mathbf{I}$ is the identity matrix. We define the $(m,n)$ element of the $N \times N$ discrete Fourier transform matrix (DFT) $\mathbf{F}$ as $F[m,n] = 1/\sqrt{N} e^{\frac{-j2\pi mn}{N}}$. The superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ designate transposition and Hermitian transposition respectively. The notation $\| \cdot \|_2$ stands for the Euclidian norm and $\delta(\cdot)$ is the Dirac delta function. For a set $\mathcal{A}$ we define the indicator function as $I_{\mathcal{A}}(x) = 1$, if $x \in \mathcal{A}$ and $I_{\mathcal{A}}(x) = 0$, otherwise. The row-wise concatenation of two matrices $\mathbf{A}, \mathbf{B}$ reads $[\mathbf{A}|\mathbf{B}]$. A function $f : \mathcal{E} \to \mathcal{F}$ maps the set $\mathcal{E}$ to the set $\mathcal{F}$ and its support is $\mathrm{supp}(f) =$

$\{x \in \mathcal{E} |\ f(x) \neq 0\}$. The convolution of two functions $f$ and $g$ is denoted by $f * g$. We use $\mathrm{CN}(\cdot|\mathbf{a}, \mathbf{B})$ to designate a multivariate complex Gaussian pdf with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{B}$, and $\mathrm{Ga}(\cdot; a, b)$ for the gamma pdf with shape and rate parameters $a$ and respectively $b$. The expected value of $f(x)$ w.r.t. the pdf $q(x)$ is $\langle f(x) \rangle_{q(x)}$.

# 2 System model

In this section we develop the signal model and the associated probabilistic model of an OFDM system transmitting across a channel with MED longer than the CP duration. Since signals transmitted in the previous signaling interval interfere with the signals received in the current signaling interval, our derivation considers all signals received over one signaling block.

## 2.1 Model of the OFDM signal received in insufficient CP conditions

During the $n$th signaling interval a vector $\mathbf{u}_n \in \{0,1\}^K$ of information bits is encoded with a code rate $R$ and interleaved into the vector $\mathbf{c}_n = \mathcal{C}(\mathbf{u}_n) = [\mathbf{c}_n^{(0)\mathrm{T}}, ..., \mathbf{c}_n^{(N_D-1)\mathrm{T}}]^{\mathrm{T}} \in \{0,1\}^{K/R}$, with $\mathcal{C}$ denoting the coding and interleaving mapping. The entries of $\mathbf{c}_n^{(d)} \in \{0,1\}^Q, d \in [0 : N_D - 1]$ are $c_n^{(d)}[q], q \in [0 : Q-1], RN_DQ = K$. The code vector $\mathbf{c}_n$ is modulated onto a vector of complex symbols that are interleaved with pilot symbols producing the symbol vector $\mathbf{x}_n \in \mathbb{C}^N$. The $k$-th entry $x_n[k]$ of $\mathbf{x}_n$ is a pilot symbol if $k \in \mathcal{P}$ and a data symbol if $k \in \mathcal{D}$, where $\mathcal{P} = \{p_0, ..., p_{N_P-1}\}$ and $\mathcal{D} = \{d_0, ..., d_{N_D-1}\}$ represent the subsets of pilot and data indices respectively so that $\mathcal{P} \cup \mathcal{D} = [0 : N-1]$, $\mathcal{P} \cap \mathcal{D} = \varnothing$. Specifically, for $k = p_j$, $x_n[k] \in \mathcal{S}_p$ and for $k = d_i$, $x_n[k] = \mathcal{M}(\mathbf{c}_n^{(i)})$. The function $\mathcal{M} : \{1,0\}^Q \to \mathcal{S}_d$ denotes the complex modulation mapping, and the sets $\mathcal{S}_p$ and $\mathcal{S}_d$ are complex symbol constellations.

The symbol vector $\mathbf{x}_n$ is passed through an inverse DFT block, yielding the vector $\mathbf{s}_n = \mathbf{F}^{\mathrm{H}}\mathbf{x}_n$. A $\mu$-sample long cyclic prefix (CP) is prepended to $\mathbf{s}_n$ to produce the vector of samples of the OFDM symbol transmitted in the $n$th signaling interval. These samples are modulated using a transmit pulse-shape filter with response $\psi_{tx}(t), \mathrm{supp}(\psi_{tx}) = [0, T]$ yielding the OFDM waveform transmitted in the $n$th signaling interval: $s_n(t) = \sum_{i=-\mu}^{N-1} s_n[i] \cdot \psi_{tx}\left(t - (i + n(\mu+N))T_s\right),\ t \in \Delta_n$, where $s_n[-i] = s_n[N-i], i \in [1 : \mu]$, $\Delta_n = \left[((n-1)\mu + nN)T_s, (n\mu + (n+1)N)T_s\right)$ and $T_s$ represents the sampling time. Alternatively, we can express $s_n(t) = (\tilde{s}_n * \psi_{tx})(t)$ where $\tilde{s}_n(t) = \sum_{i=-\mu}^{N-1} s_n[i]\delta(t - (i + n(\mu+N))T_s)$.

The signal transmitted across the wireless channel consists of a block of $B$ consecutive OFDM waveforms: $s(t) = \sum_{n=0}^{n=B-1} s_n(t)$. The channel impulse

response (CIR) is considered invariant over one signaling block[2]. The CIR exhibits $\bar{L}$ multipath components, i.e.

$$\bar{g}(t,\tau) = \bar{g}(\tau) = \sum_{l=0}^{\bar{L}-1} \bar{\alpha}[l]\delta(\tau - \bar{\tau}[l]), \quad t \in \Delta_0 \cup ... \cup \Delta_{B-1}. \tag{A.1}$$

It is entirely characterized by the number of multipath components $\bar{L}$, complex gains in $\bar{\boldsymbol{\alpha}} \in \mathbb{C}^{\bar{L}}$ and delays[3] in $\bar{\boldsymbol{\tau}} \in \mathbb{R}^{\bar{L}}$. Contrary to classical signal models, we allow the delays to exceed the duration of the CP, $\mu Ts$. We restrain however our analysis to responses with a maximum excess delay no longer than the duration of the OFDM symbol, i.e. $\bar{\tau}[L-1] \leq (\mu+N)T_s$.

The received signal is the convolution of the transmitted signal and the CIR, corrupted by additive white Gaussian noise (AWGN): $s(t) * \bar{g}(t,\tau) + w(t) = \sum_{n=0}^{n=B-1} \tilde{s}_n(t) * \psi_{tx}(t) * \bar{g}(t,\tau) + w(t)$. The receive filter with response $\psi_{rx}(t)$, $\mathrm{supp}(\psi_{rx}) = [0,T]$, is applied to the received signal to yield the signal $r(t) = \sum_{n=0}^{n=B-1} \tilde{s}_n(t) * \bar{g}(t,\tau) * \phi(t) + \bar{w}(t)$, where $\phi(t) = (\psi_{tx} * \psi_{rx})(t), \mathrm{supp}(\phi) = [0,2T]$ and $\bar{w}(t) = (w * \psi_{rx})(t)$. Since $\bar{\tau}[L-1] \leq (\mu+N)T_s$ the restriction of $r(t)$ to the $n$th signaling interval $\Delta_n$ contains the contributions of the $n$th and $(n-1)$th transmit OFDM waveforms:

$$\begin{aligned} r(t) = \sum_{m=n-1}^{n} \sum_{i=-\mu}^{N-1} s_m[i] \sum_{l=0}^{\bar{L}-1} \bar{\alpha}[l]\phi\left(t - \bar{\tau}[l] - (i + m(\mu+N))T_s\right) \\ + \bar{w}(t), \quad t \in \Delta_n. \end{aligned} \tag{A.2}$$

The signal in (A.2) is sampled with rate $(T_s)^{-1}$. For each $n \in [0 : B-1]$, the samples at time instants in $\Delta_n$ corresponding to the CP are discarded while the remaining samples are collected in the vector of observations $\mathbf{r}_n$ with entries $r_n[k] = r((k + n(\mu+N))T_s), k \in [0 : N-1]$.

We define the composite CIR during the $n$th signaling interval to be $\bar{q}(\tau) = \sum_{l=0}^{\bar{L}-1} \bar{\alpha}[l]\phi\left(\tau - \bar{\tau}[l]\right)$ with $\mathrm{supp}(\bar{q}) = [\bar{\tau}[0],\ \bar{\tau}[L-1] + 2T]$. This enables us to express the $k$th entry, $k \in [0 : N-1]$, of $\mathbf{r}_n$ as $r_n[k] = \sum_{i=-\mu}^{N-1} s_{n-1}[i]\bar{q}((k-i+(\mu+N))T_s) + \sum_{j=-\mu}^{N-1} s_n[j]\bar{q}((k-j)T_s) + \bar{w}_n[k]$, where $\bar{w}_n[k] = \bar{w}((k+n(\mu+N))T_s)$ is the $k$th entry of the noise vector $\bar{\mathbf{w}}_n$.

The vector $\mathbf{r}_n$ is passed through a DFT block to yield $\mathbf{y}_n = \mathbf{F}\mathbf{r}_n$, which equivalently reads:

$$\mathbf{y}_n = \bar{\mathbf{H}}\mathbf{x}_n + \mathbf{F}\bar{\mathbf{C}}\mathbf{F}^{\mathrm{H}}\mathbf{x}_n + \mathbf{F}\bar{\mathbf{S}}\mathbf{F}^{\mathrm{H}}\mathbf{x}_{n-1} + \boldsymbol{\xi}_n, \quad n = [0 : B-1] \tag{A.3}$$

[2]We consider the OFDM communication system to work with short blocks over which the CIR can be treated as static.

[3]We assume, without loss of generality, that the entries of the delay vector $\bar{\boldsymbol{\tau}}$ are sorted in ascending order, i.e., $\bar{\tau}[l] < \bar{\tau}[l+1], l \in [0 : \bar{L}-1]$.

where $\boldsymbol{\xi}_n = \mathbf{F}\bar{\mathbf{w}}_n$. The diagonal matrix $\bar{\mathbf{H}}$ and the matrices $\bar{\mathbf{C}}$ and $\bar{\mathbf{S}}$ have entries

$$\bar{H}[i,i] = \sum_{v=0}^{\bar{M}-1} \sqrt{N}\bar{q}(vT_s)F[i,v] \tag{A.4a}$$

$$\bar{C}[m,i] = -\bar{q}((N+m-i)T_s)I_{[0:\bar{E}-1]}(m)I_{[m+(N-\bar{E}-\mu):N-\mu-1]}(i) \tag{A.4b}$$

$$\bar{S}[m,i] = \bar{q}((N+\mu+m-i)Ts)I_{[0:\bar{E}-1]}(m)I_{[m+(N-\bar{E}):N-1]}(i) \tag{A.4c}$$

where $m,i \in [0:N-1]$. We define $\bar{M} = \lceil(\bar{\tau}[\bar{L}-1]+2T)/T_s\rceil$ and $\bar{E} = \bar{M}-\mu$.

We observe in (A.3) the explicit ICI and ISI contributions in the received signal through the terms containing the matrices $\bar{\mathbf{C}}$ and $\bar{\mathbf{S}}$ respectively.

The dependency of $\mathbf{y}_n$ on the CIR can be made more explicit by reformulating (A.3) as

$$\mathbf{y}_n = \bar{\mathbf{A}}_n\bar{\boldsymbol{\alpha}} + \boldsymbol{\xi}_n, \quad n = [0:B-1] \tag{A.5}$$

where $\bar{\mathbf{A}}_n = \bar{\mathbf{V}}_n\bar{\boldsymbol{\Phi}}$ and $\bar{\mathbf{V}}_n = \mathbf{X}_n\sqrt{N}\mathbf{F}_{N\times\bar{M}} + \bar{\boldsymbol{\Xi}}_n$.

The entries of the matrix $\bar{\boldsymbol{\Phi}} \in \mathbb{C}^{\bar{M}\times\bar{L}}$ are $\bar{\Phi}[m,l] = \phi(mT_s - \bar{\tau}[l])$, $m \in [0:\bar{M}-1], l \in [0:\bar{L}-1]$ and the matrix $\mathbf{X}_n \in \mathbb{C}^{N\times N}$ is diagonal with diagonal entries $X_n[i,i] = x_n[i], i \in [0:N-1]$. We define $\bar{\boldsymbol{\Xi}}_n = \left[\mathbf{0}_{N\times\mu}|\bar{\mathbf{Y}}_n\right] \in \mathbb{C}^{N\times\bar{M}}$. The rows of the matrix $\bar{\mathbf{Y}}_n \in \mathbb{C}^{N\times\bar{E}}$ are $\bar{\mathbf{Y}}_n[k,\cdot] = \frac{1}{N}\left(\mathbf{x}_{n-1}^{\mathrm{T}}\boldsymbol{\varpi} - \mathbf{x}_n^{\mathrm{T}}\right)\bar{\boldsymbol{\Lambda}}^{(k)}, k \in [0:N-1]$. The matrix $\boldsymbol{\varpi} \in \mathbb{C}^{N\times N}$ is diagonal with diagonal entries $\varpi[i,i] = \omega^{\mu i}, i \in [0:N-1]$. The matrix $\bar{\boldsymbol{\Lambda}}^{(k)} \in \mathbb{C}^{N\times\bar{E}}$ has entries $\bar{\Lambda}^{(k)}[a,b] = \sum_{u=0}^{\bar{E}-1}\chi(u,b+\mu,k,a), a \in [0:N-1], b \in [0:\bar{E}-1]$ with $\chi(u,r,k,a) = \omega^{u(k-a)+ra}I_{[u+\mu+1,\bar{M}-1]}(r)$.

## 2.2 Proposed estimation model

From the signal model presented in (A.3) and its equivalent representation (A.5) one can observe that in order to decode the transmitted message $\mathbf{u}_n$, $n = [0:B-1]$, the receiver needs to deal with a set of unknown random variables which are not, in principle, of direct interest to the receiver. These variables, usually referred to as nuisance parameters, comprise the entries in the vector of channel weights $\bar{\boldsymbol{\alpha}}$, and the vector of multipath delays $\bar{\boldsymbol{\tau}}$, the component variance of the noise vector $\boldsymbol{\xi}_n$, and the symbols in $\mathbf{x}_n, n \in [0:B-1]$. One possible approach to design the receiver is to attempt the joint estimation of all the unknown variables in the model. Such strategy leads, however, to severe tractability problems, mainly due to the difficulty in estimating the vector of multipath delays $\bar{\boldsymbol{\tau}}$. To overcome this difficulty, we adopt an SBI approach for channel estimation. SBI techniques for channel estimation are founded on the assumption that the CIR is sparse, i.e. composed of a few dominant multipath components. Since the delays of these multipath components are unknown at

the reception, the CIR during the $n$th signaling interval is approximated as the sum of a large number of multipath components whose delays are evenly spaced within a given delay range, i.e.

$$g(t,\tau) = g(\tau) = \sum_{l=0}^{L-1} \alpha[l]\delta(\tau - \tau_s[l]), \quad t \in \Delta_0 \cup ... \cup \Delta_{B-1} \tag{A.6}$$

Here we define the postulated vector of channel weights $\boldsymbol{\alpha} \in \mathbb{C}^L$ with entries $\alpha[l]$ from (A.6); the delay vector with delay resolution $\Delta\tau$ is $\boldsymbol{\tau}_s \in \mathbb{R}^L$ with entries $\tau_s[p] = p\Delta\tau, p \in [0 : L-1]$. We select $L$ such that $L \gg \bar{L}$. Given the approximation (A.6) in conjunction with the aforementioned sparsity assumption, we expect $\boldsymbol{\alpha}$ to be sparse, i.e. many of its entries to be zero. Exploiting this property, the SBI framework enforces sparse estimates of $\boldsymbol{\alpha}$ by employing a probabilistic model that assigns a sparsity-inducing prior distribution to this vector.

Following the above rationale, we formulate a new (approximate) estimation model by applying the approximation (A.6) to the signal model (A.3) and (A.5). To that end, we define the approximate dictionary matrix $\mathbf{A}_n \in \mathbb{C}^{N\times L}$ with entries

$$A_n[a,b] = \sum_{k=0}^{M-1} V_n[a,k]\Phi[k,b], \quad a \in [0 : N-1],\ b \in [0 : L-1], \tag{A.7}$$

where $M = \lceil (\tau_s[L-1] + 2T)/T_s \rceil$ and $\Phi[k,b] = \phi(kT_s - \tau_s[b])$. The matrix $\mathbf{V}_n$ is defined analogously to $\bar{\mathbf{V}}_n$ by replacing $\bar{M}$ with $M$. We obtain thus the following approximations of (A.5) and (A.3):

$$\mathbf{y}_n = \mathbf{A}_n\boldsymbol{\alpha} + \boldsymbol{\xi}_n, \tag{A.8a}$$

$$\mathbf{y}_n = \mathbf{H}\mathbf{x}_n + \mathbf{F}\mathbf{C}\mathbf{F}^{\mathrm{H}}\mathbf{x}_n + \mathbf{F}\mathbf{S}\mathbf{F}^{\mathrm{H}}\mathbf{x}_{n-1} + \boldsymbol{\xi}_n, \quad n = [0 : B-1]. \tag{A.8b}$$

The noise vector $\boldsymbol{\xi}_n$ is white and Gaussian, i.e. $p(\boldsymbol{\xi}_n) = \mathrm{CN}(\boldsymbol{\xi}_n; \mathbf{0}, \lambda^{-1}\mathbf{I}_N)$. The matrices $\mathbf{H}, \mathbf{C}, \mathbf{S}$ are also defined analogously to $\bar{\mathbf{H}}, \bar{\mathbf{C}}, \bar{\mathbf{S}}$ in (3.1) by replacing $\bar{M}$ with $M$, and $\bar{q}$ with $q(\tau) = \sum_{l=0}^{L-1} \alpha[l]\phi(\tau - \tau_s[l])$.

Using (A.8a), we can express the joint pdf of all unknown random variables of the system as

$$\begin{aligned} &p(\mathbf{u}_0, ..., \mathbf{u}_{B-1}, \mathbf{c}_0, ..., \mathbf{c}_{B-1}, \mathbf{x}_0, ..., \mathbf{x}_{B-1}, \lambda, \boldsymbol{\alpha}, \mathbf{y}_0, ..., \mathbf{y}_{B-1}) \propto \\ &p(\lambda)p(\boldsymbol{\alpha}) \prod_{n=0}^{B-1} p(\mathbf{y}_n|\mathbf{x}_n, \mathbf{x}_{n-1}, \lambda, \boldsymbol{\alpha}) I_{\{\mathcal{M}(\mathbf{c}_n)\}}(\mathbf{x}_n) I_{\{\mathcal{C}(\mathbf{u}_n)\}}(\mathbf{c}_n) p(\mathbf{u}_n) \end{aligned} \tag{A.9}$$

where $p(\mathbf{y}_n|\mathbf{x}_n, \mathbf{x}_{n-1}, \lambda, \boldsymbol{\alpha}) = \mathrm{CN}(\mathbf{y}_n; \mathbf{A}_n\boldsymbol{\alpha}, \lambda^{-1}\mathbf{I}_N)$ and $\mathbf{x}_{-1} = \mathbf{0}$.

The factors $I_{\{\mathcal{M}(\mathbf{c}_n)\}}(\mathbf{x}_n) = \prod_{k=0}^{N_D-1} I_{\{\mathcal{M}(\mathbf{c}_n^{(k)})\}}(x_n[d_k])$ and $I_{\{\mathcal{C}(\mathbf{u}_n)\}}(\mathbf{c}_n)$ enforce modulation and coding constraints. We assume that the information

bits are generated using a binary symmetric source, i.e. $p(\mathbf{u}_n) = \prod_{v=0}^{K-1} p(u_n[v])$, with $p(u_n[v]) = \frac{1}{2}$, $u_n[v] \in \{0, 1\}$. We select the noise precision prior to be the non-informative, i.e. $p(\lambda) \propto 1/\lambda$. To enforce sparsity on the estimator of $\boldsymbol{\alpha}$ we use a sparsity-inducing probabilistic modeling of the prior pdf $p(\boldsymbol{\alpha})$. We model the entries in $\boldsymbol{\alpha}$ as Gaussian scale mixtures, i.e. $p(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = p(\boldsymbol{\alpha}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})$ with $p(\boldsymbol{\alpha}|\boldsymbol{\gamma}) = \mathrm{CN}(\boldsymbol{\alpha}; \mathbf{0}, \boldsymbol{\Gamma})$ and $\boldsymbol{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$. Following the approach in [28, 40], we select the mixing density as $p(\boldsymbol{\gamma}) = \prod_{p=0}^{L-1} \mathrm{Ga}(\gamma[p]; \epsilon, \eta)$.

# 3 Message passing receiver design for insufficient CP conditions

Given the estimation model (A.8) and its probabilistic description (A.9), the task of the receiver is to retrieve the transmitted bit vectors $\mathbf{u}_n, n \in [0 : B-1]$. The optimum receiver computes $p(u_n[k]|\mathbf{y}_n, n \in [0 : B-1])$ for each $n \in [0 : B-1], k \in [0 : K-1]$. The high complexity of the involved marginalization procedure prevents any feasible implementation of this receiver. Instead we resort to variational Bayesian inference techniques, which estimate the posterior probability distributions of each variable in the probabilistic model, i.e. the information bits, the channel weights, the noise variance, and the transmitted symbols. Specifically, we design a message-passing receiver algorithm that implements the approximate inference technique [37] applied to the factor graph representation [41] of (A.9). In the design we will use both representations (A.8a) and (A.8b) of the signal model. The representation (A.8a) is suitable for deriving estimators of the weight vector $\boldsymbol{\alpha}$. Furthermore, since this vector consists of few non-zero elements, finding such estimators corresponds to solving a sparse Bayesian learning problem. The representation (A.8b) is tailored for retrieving the data symbols $\mathbf{x}_n$, which is done by solving a classical equalization problem. However, given that $\mathbf{A}_n$ in (A.8a) depends on $\mathbf{x}_n, \mathbf{x}_{n-1}$ and $\mathbf{C}, \mathbf{S}, \mathbf{H}$ in (A.8b) depend on $\boldsymbol{\alpha}$, the two problems cannot be treated separately.

## 3.1 MF-BP message-passing technique: preliminaries

Consider a probabilistic model consisting of a set $\mathcal{Z}$ of random variables and characterized by a joint pdf that factorizes according to $p(\mathbf{z}) \propto \prod_{f_a \in \mathcal{F}} f_a(\mathbf{z}_a)$, where $\mathbf{z}$ is a vector containing all random variables $z_i \in \mathcal{Z}$ and $\mathbf{z}_a$ is a vector with all variables $z_i$ that are argument of the local function $f_a$ belonging to the set of functions $\mathcal{F}$. The (bipartite) factor graph of the probabilistic model[4] specified by $p(\mathbf{z})$ consists of a variable node for each variable $z_i$, a factor node for each factor $f_a$, and edges which connect a variables node $z_i$ to a factor node

[4]Factor graphs provide graphical representations of the statistical dependencies of the variables in a probabilistic model. Note that a probabilistic model can have different (equivalent) graphical representation.

$f_a$ if, and only if, $z_i$ is an argument of $f_a$ [41]. Given the above probabilistic model, a typical inference problem is computing the marginal pdfs of the variables of interest, i.e. $p(z_i)$. Computing the marginals, however, is often infeasible - especially for large probabilistic models with complex dependencies - and approximate inference approaches are sought instead. Approximate inference techniques work with beliefs - i.e. approximate pdfs - $q(z_i)$ of the marginals, such that $q(z_i) \approx p(z_i)$. The computation of the beliefs is often iterative and can typically be formulated as a message-passing algorithm in the factor graph associated to the probabilistic model. Messages are iteratively exchanged along the edges of the graph, according to some specified updating rules, called scheduling, until convergence is reached. Eventually the beliefs $q(z_i)$ are retrieved.

The MF-BP framework [37] divides the set $\mathcal{F}$ of factor nodes into two disjoint sets $\mathcal{F}_{\text{BP}}$ and $\mathcal{F}_{\text{MF}}$, i.e. $\mathcal{F}_{\text{BP}} \cup \mathcal{F}_{\text{MF}} = \mathcal{F}, \quad \mathcal{F}_{\text{BP}} \cap \mathcal{F}_{\text{MF}} = \emptyset$. Denoting the message passed from the factor node $f_a$ to the variable node $z_i \in \mathbf{z}_a$ as $m_{f_a \to z_i}$ and the message exchanged in the opposite direction as $n_{z_i \to f_a}$, the rules [37] for updating the messages and the beliefs $q(z_i)$ are defined in Table A.1. By abuse

**Table A.1:** MF-BP rules

| | |
|---|---|
| $m^{\text{MF}}_{f_a \to z_i}(z_i) \propto \exp\left(\sum_{\mathbf{z}_a \setminus z_i} \log f_a(\mathbf{z}_a) \prod_{j: z_j \in \mathcal{N}(f_a) \setminus z_i} n_{z_j \to f_a}(z_j)\right),$ | |
| $f_a \in \mathcal{F}_{\text{MF}}, z_j \in \mathcal{N}(f_a)$ | (I) |
| $m^{\text{BP}}_{f_a \to z_i}(z_i) \propto \sum_{\mathbf{z}_a \setminus z_i} f_a(\mathbf{z}_a) \prod_{j: z_j \in \mathcal{N}(f_a) \setminus z_i} n_{z_j \to f_a}(z_j),$ | |
| $f_a \in \mathcal{F}_{\text{BP}}, z_j \in \mathcal{N}(f_a)$ | (II) |
| $n_{z_i \to f_a}(z_i) \propto \prod_{f_c \in \mathcal{N}(z_i) \cap \mathcal{F}_{\text{MF}}} m^{\text{MF}}_{f_c \to z_i}(z_i) \prod_{f_c \in \mathcal{N}(z_i) \cap \mathcal{F}_{\text{BP}} \setminus f_a} m^{\text{BP}}_{f_c \to z_i}(z_i),$ | |
| $z_i \in \mathcal{Z}, f_a \in \mathcal{N}(z_i)$ | (III) |
| $q(z_i) \propto \prod_{f_c \in \mathcal{N}(z_i) \cap \mathcal{F}_{\text{MF}}} m^{\text{MF}}_{f_c \to z_i}(z_i) \prod_{f_c \in \mathcal{N}(z_i) \cap \mathcal{F}_{\text{BP}}} m^{\text{BP}}_{f_c \to z_i}(z_i), z_i \in \mathcal{Z}$ | (IV) |
| $\mathcal{N}(z_i) \subset \mathcal{F}$ is the subset of functions $f_a$ that have variable $z_i$ as an argument. | |
| $\mathcal{N}(f_a) \subset \mathcal{Z}$ is the subset of variables that are arguments of $f_a$. | |

of terminology, we refer to messages of the form (I) and (II) as respectively MF and BP messages. Note that the MF-BP framework is presented above under the assumption that all variables $z_i \in \mathcal{Z}$ are scalar variables and, hence, the computed beliefs represent approximations of marginal pdfs of scalar variables. In practice, it is possible to group variables into a single vector variable, say $\boldsymbol{z}_g$ with entries $z_i \in \mathcal{Z}_g$ for some $\mathcal{Z}_g \subseteq \mathcal{Z}$. In this case, the vector-valued variable is represented in the factor graph as a single node. The message computation is formally extended to apply to such vector-valued variable, and the algorithm computes a joint belief $q(\boldsymbol{z}_g)$ approximating the joint marginal pdf $p(\boldsymbol{z}_g)$. The selected grouping of variables reflects an assumption on the constraints that the beliefs should fulfill, as discussed in [37]. In the remainder of this section we will explore two different grouping options - and, consequently, two

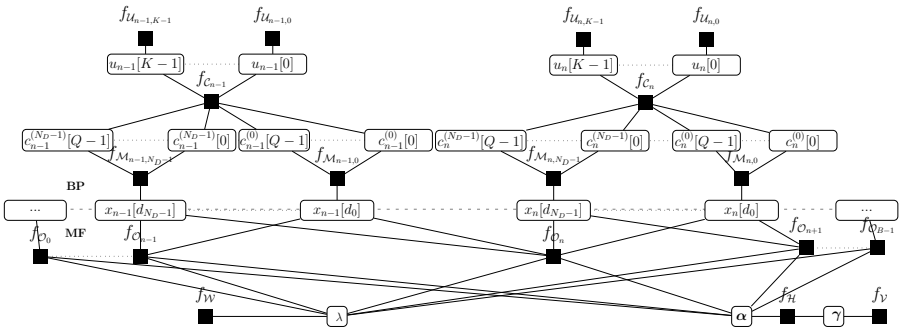


**Fig. A.1:** Factor graph representation of the probabilistic model (A.10).

different factor graphs and two different message-passing algorithms - for the probabilistic model (A.9).

## 3.2 Computation of messages and beliefs

In the following, we apply the MF-BP framework to the probabilistic model presented in Section II. First, we express the joint pdf (A.9) as

$$\begin{aligned} &p(\mathbf{u}_0, ..., \mathbf{u}_{B-1}, \mathbf{c}_0, ..., \mathbf{c}_{B-1}, \mathbf{x}_0, ..., \mathbf{x}_{B-1}, \lambda, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \mathbf{y}_0, ..., \mathbf{y}_{B-1}) = \\ &f_{\mathcal{W}}(\lambda) f_{\mathcal{H}}(\boldsymbol{\alpha}, \boldsymbol{\gamma}) f_{\mathcal{V}}(\boldsymbol{\gamma}) \prod_{n=0}^{B-1} f_{\mathcal{O}_n}(\mathbf{x}_n, \mathbf{x}_{n-1}, \lambda, \boldsymbol{\alpha}) f_{\mathcal{C}_n}(\mathbf{c}_n, \mathbf{u}_n) \cdot \\ &\prod_{k=0}^{N_D-1} f_{\mathcal{M}_{n,k}}(x_n[d_k], \mathbf{c}_n^{(k)}) \prod_{v=0}^{K-1} f_{\mathcal{U}_{n,v}}(u_n[v]), \end{aligned} \tag{A.10}$$

where $f_{\mathcal{O}_n}(\mathbf{x}_n, \mathbf{x}_{n-1}, \lambda, \boldsymbol{\alpha}) = p(\mathbf{y}_n|\mathbf{x}_n, \mathbf{x}_{n-1}, \lambda, \boldsymbol{\alpha})$, $f_{\mathcal{W}}(\lambda) = p(\lambda)$, $f_{\mathcal{H}}(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = p(\boldsymbol{\alpha}|\boldsymbol{\gamma})$, $f_{\mathcal{V}}(\boldsymbol{\gamma}) = p(\boldsymbol{\gamma})$, $f_{\mathcal{M}_{n,k}}(x_n[d_k], \mathbf{c}_n^{(k)}) = I_{\{\mathcal{M}(\mathbf{c}_n^{(k)})\}}(x_n[d_k])$, $f_{\mathcal{C}_n}(\mathbf{c}_n, \mathbf{u}_n) = I_{\{\mathcal{C}(\mathbf{u}_n)\}}(\mathbf{c}_n)$, and $f_{\mathcal{U}_{n,v}}(u_n[v]) = p(u_n[v]) = 1/2$. Based on (A.10), we cast the factor graph depicted in Fig. A.1.

We further define $\mathcal{F}_{\text{MF}} = \{f_{\mathcal{W}}, f_{\mathcal{V}}, f_{\mathcal{H}}, f_{\mathcal{O}_n} | n = [0 : B-1]\}$ and $\mathcal{F}_{\text{BP}} = \mathcal{F} \backslash \mathcal{F}_{\text{MF}}$. Henceforth, we will refer to the subgraph corresponding to the factorization $\prod_{f_a \in \mathcal{F}_{\text{BP}}} f_a(\mathbf{x}_a)$ as the BP part, and similarly to that representing $\prod_{f_a \in \mathcal{F}_{\text{MF}}} f_a(\mathbf{x}_a)$ as the MF part. We chose to treat the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ as single variables, each represented by a single variable node. An alternative representation with individual (separate) variable nodes for each of the entries of $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ will be discussed later in this section.

An analogy can be made between passing of messages along edges of the graph and the typical operations of an OFDM receiver, as we will see in

the following. Propagating messages from the node $f_{\mathcal{O}_n}$, to all factor nodes $f_{\mathcal{U}_{n,k}}, k \in [0:K-1]$ and back (i.e. exchanging messages in the (upper) BP part of the graph) corresponds to soft[5] demapping, soft decoding, deinterleaving, interleaving, soft coding and soft mapping. Computing the beliefs of $\lambda, \boldsymbol{\alpha}, \boldsymbol{\gamma}$ in the (lower) MF part of the graph corresponds to channel estimation.

**Computation of the beliefs of the data symbols**

The information inferred on $\mathbf{x}_n$ from the observation vectors $\mathbf{y}_n$, together with estimates of the channel weights and the noise precision, are propagated upwards in the graph through the set of messages $m^{\mathrm{MF}}_{\mathcal{O}_n \to x_n[d_k]}$, $n \in [0 : B-1]$ . Subsequently, the messages $n_{x_n[d_k]\to f_{\mathcal{M}_{n,k}}}$, $m^{\mathrm{BP}}_{f_{\mathcal{M}_{n,k}}\to c_n^k[m]}$, $n_{c_n^k[m]\to f_{\mathcal{C}_n}}$, $m^{\mathrm{BP}}_{f_{\mathcal{C}_n}\to u_n[a]}$, $n_{u_n[a]\to f_{\mathcal{U}_{n,a}}}$ $k \in [0:N_D-1], m \in [0:Q-1], a \in [0:K-1]$, corresponding to the operations of soft demapping and decoding, are computed. Similarly, propagating the messages downwards, i.e. $m^{\mathrm{BP}}_{f_{\mathcal{U}_{n,a}}\to u_n[a]}$, $n_{u_n[a]\to f_{\mathcal{C}_n}}$, $m^{\mathrm{BP}}_{f_{\mathcal{C}_n}\to c_n^k[m]}$, $n_{c_n^k[m]\to f_{\mathcal{M}_{n,k}}}$, $m^{\mathrm{BP}}_{f_{\mathcal{M}_{n,k}}\to x_n[d_k]}$, is equivalent to soft encoding and re-mapping. It is well known that propagation of BP messages through nodes $f_{\mathcal{M}_{n,k}}$ and $f_{\mathcal{C}_n}$ corresponds to classical demapping [42] and decoding algorithms for most channel codes (e.g. the BCJR algorithm for convolutional codes [41], or the turbo-decoding algorithm for turbo codes [43, 44]). We now focus on the computation of the beliefs of the data symbols, which read

$$\begin{aligned} q(x_n[d_k]) \propto & m^{\mathrm{MF}}_{f_{\mathcal{O}_n}\to x_n[d_k]}(x_n[d_k])\; m^{\mathrm{MF}}_{f_{\mathcal{O}_{n+1}}\to x_n[d_k]}(x_n[d_k]) \\ & m^{\mathrm{BP}}_{f_{\mathcal{M}_{n,k}}\to x_n[d_k]}(x_n[d_k]). \end{aligned} \tag{A.11}$$

Using (A.8b), the first two factors in (A.11) are given by

$$\begin{aligned} & m^{\mathrm{MF}}_{f_{\mathcal{O}_n}\to x_n[d_k]}(x_n[d_k]) \propto \\ & \exp\left( - \langle \lambda \rangle_{q(\lambda)} \left\langle \| \mathbf{y}_n - \mathbf{M}\mathbf{x}_n - \mathbf{N}\mathbf{x}_{n-1} \|_2^2 \right\rangle_{q(\mathbf{x}_{n\setminus d_k})q(\mathbf{x}_{n-1})q(\boldsymbol{\alpha})} \right), \end{aligned} \tag{A.12}$$

where the entries of $\mathbf{x}_{n\setminus d_k}$ are $x_{n\setminus d_k}[d] = x_n[d]$, if $d \neq d_k$ and $x_{n\setminus d_k}[d] = 0$, if $d = d_k$, $\mathbf{M} = \mathbf{H} + \mathbf{F}\mathbf{C}\mathbf{F}^{\mathrm{H}}$ and $\mathbf{N} = \mathbf{F}\mathbf{S}\mathbf{F}^{\mathrm{H}}$, and

$$\begin{aligned} & m^{\mathrm{MF}}_{f_{\mathcal{O}_{n+1}}\to x_n[d_k]}(x_n[d_k]) \propto \\ & \exp\left( - \langle \lambda \rangle_{q(\lambda)} \left\langle \| \mathbf{y}_{n+1} - \mathbf{M}\mathbf{x}_{n+1} - \mathbf{N}\mathbf{x}_n \|_2^2 \right\rangle_{q(\mathbf{x}_{n\setminus d_k})q(\mathbf{x}_{n+1})q(\boldsymbol{\alpha})} \right). \end{aligned} \tag{A.13}$$

[5] Henceforth, we use the terminology "soft operations" to designate operations that produce soft values of a model parameter, i.e. compute the first and second moments of said parameter with respect to its (current) belief.

The product of messages (A.12) and (A.13) is the message $n_{x_n[d_k]\to f_{\mathcal{M}_{n,k}}}$ which can be recast as $n_{x_n[d_k]\to f_{\mathcal{M}_{n,k}}}(x_n[d_k]) \propto \mathrm{CN}(x_n[d_k]; \mu_{x_n[d_k]}, \sigma^2_{x_n[d_k]})$ with

$$\left(\sigma^2_{x_n[d_k]}\right)^{-1} = \langle\lambda\rangle_{q(\lambda)} \left\langle \| \mathbf{M}[\cdot, d_k] \|_2^2 \right\rangle_{q(\boldsymbol{\alpha})} + \langle\lambda\rangle_{q(\lambda)} \left\langle \| \mathbf{N}[\cdot, d_k] \|_2^2 \right\rangle_{q(\boldsymbol{\alpha})} \tag{A.14}$$

$$\begin{aligned}\mu_{x_n[d_k]} \left(\sigma^2_{x_n[d_k]}\right)^{-1} = \langle\lambda\rangle_{q(\lambda)} \Big( & \left\langle \mathbf{M}^{\mathrm{H}}[\cdot, d_k] \right\rangle_{q(\boldsymbol{\alpha})} \cdot \\ & \left\langle \mathbf{y}_n - \mathbf{M}\mathbf{x}_{n\setminus d_k} - \mathbf{N}\mathbf{x}_{n-1} \right\rangle_{q(\mathbf{x}_{n\setminus d_k})q(\mathbf{x}_{n-1})q(\boldsymbol{\alpha})} + \\ & \left\langle \mathbf{N}^{\mathrm{H}}[\cdot, d_k] \right\rangle_{q(\boldsymbol{\alpha})} \left\langle \mathbf{y}_{n+1} - \mathbf{N}\mathbf{x}_{n\setminus d_k} - \mathbf{M}\mathbf{x}_{n+1} \right\rangle_{q(\mathbf{x}_{n\setminus d_k})q(\mathbf{x}_{n+1})q(\boldsymbol{\alpha})} \Big).\end{aligned} \tag{A.15}$$

The computation of $\langle \mathbf{M} \rangle_{q(\boldsymbol{\alpha})}, \langle \mathbf{N} \rangle_{q(\boldsymbol{\alpha})}, \left\langle \mathbf{M}^{\mathrm{H}}\mathbf{M} \right\rangle_{q(\boldsymbol{\alpha})}$ and $\left\langle \mathbf{N}^{\mathrm{H}}\mathbf{N} \right\rangle_{q(\boldsymbol{\alpha})}$ is provided in Appendix B. The BP message in (A.11) reads

$$\begin{aligned} m^{\mathrm{BP}}_{f_{\mathcal{M}_{n,k}}\to x_n[d_k]}(x_n[d_k]) \propto & \sum_{\substack{\mathbf{c}_n^{(k)} \in \{0,1\}^Q \\ Q-1}} f_{\mathcal{M}_{n,k}}(x_n[d_k], \mathbf{c}_n^{(k)}) \\ & \prod_{v=0} n_{c_n^{(k)}[v]\to f_{\mathcal{M}_{n,k}}}(c_n^{(k)}[v]). \end{aligned} \tag{A.16}$$

Once $q(x_n[d_k])$ in (A.11) is computed, $\langle \mathbf{A}_n \rangle_{q(\mathbf{x}_{n-1})q(\mathbf{x}_n)}, \left\langle \mathbf{A}_n^{\mathrm{H}}\mathbf{A}_n \right\rangle_{q(\mathbf{x}_{n-1})q(\mathbf{x}_n)}$, $n \in [0 : B-1]$ can be updated – see Appendix B.

**Computation of the beliefs of the channel weights and the noise precision**

The belief of the noise precision reads $q(\lambda) \propto m^{\mathrm{MF}}_{f_{\mathcal{W}}\to\lambda}(\lambda) \prod_{n=0}^{B-1} m^{\mathrm{MF}}_{f_{\mathcal{O}_n}\to\lambda}(\lambda)$, with $\prod_{n=0}^{B-1} m^{\mathrm{MF}}_{f_{\mathcal{O}_n}\to\lambda} \propto \lambda^{NB} \exp\left(-\lambda \sum_{n=0}^{B-1} \left\langle \| \mathbf{y}_n - \mathbf{A}_n\boldsymbol{\alpha} \|_2^2 \right\rangle_{q(\mathbf{x}_n),q(\mathbf{x}_{n-1}),q(\boldsymbol{\alpha})}\right)$ and $m^{\mathrm{MF}}_{f_{\mathcal{W}}\to\lambda}(\lambda) \propto 1/\lambda$. Subsequently, we obtain $q(\lambda) \propto \mathrm{Ga}(\lambda; NB, \sum_{n=0}^{B-1} \left\langle \| \mathbf{y}_n - \mathbf{A}_n\boldsymbol{\alpha} \|_2^2 \right\rangle_{q(\mathbf{x}_n),q(\mathbf{x}_{n-1}),q(\boldsymbol{\alpha})})$. The first moment of $q(\lambda)$ can be expressed in closed form:

$$\langle\lambda\rangle_{q(\lambda)} = \frac{NB}{\sum_{n=0}^{B-1} \left\langle \| \mathbf{y}_n - \mathbf{A}_n\boldsymbol{\alpha} \|_2^2 \right\rangle_{q(\mathbf{x}_n),q(\mathbf{x}_{n-1}),q(\boldsymbol{\alpha})}}. \tag{A.17}$$

For the computation of $\langle \mathbf{A}_n \rangle_{q(\mathbf{x}_{n-1})q(\mathbf{x}_n)}$ and $\left\langle \mathbf{A}_n^{\mathrm{H}}\mathbf{A}_n \right\rangle_{q(\mathbf{x}_{n-1})q(\mathbf{x}_n)}$, we refer to Appendix B. We propose two approaches for the estimation of the channel weights: in the first one, all weights are jointly estimated, i.e. the belief $q(\boldsymbol{\alpha})$ is computed; in the second approach, the beliefs of the individual weights $q(\alpha[p]), p \in [0 : L-1]$ are computed by assuming a fully factorized belief, i.e. $q(\boldsymbol{\alpha}) = \prod_{p=0}^{L-1} q(\alpha[p])$.

**Joint update of the channel weights**

The factor graph in Fig. A.1 represents the probabilistic model tailored to the first approach; as a result, the vector of channel weights $\boldsymbol{\alpha}$ and the vector $\boldsymbol{\gamma}$ of their prior parameters are represented each by one variable node. The belief of $\boldsymbol{\gamma}$ is the product of two messages: $q(\boldsymbol{\gamma}) \propto m^{\mathrm{MF}}_{f_{\mathcal{V}}\to\boldsymbol{\gamma}}(\boldsymbol{\gamma})\; m^{\mathrm{MF}}_{f_{\mathcal{H}}\to\boldsymbol{\gamma}}(\boldsymbol{\gamma})$ with $m^{\mathrm{MF}}_{f_{\mathcal{V}}\to\boldsymbol{\gamma}} \propto \prod_{p=0}^{L-1}\gamma[p]^{\epsilon-1}\exp\left(-\eta\gamma[p]\right)$ and $m^{\mathrm{MF}}_{f_{\mathcal{H}}\to\boldsymbol{\gamma}}(\boldsymbol{\gamma}) \propto \prod_{p=0}^{L-1}\gamma[p]^{-1}$ $\exp\left(-\boldsymbol{\gamma}[p]^{-1}\left\langle|\alpha[p]|^2\right\rangle_{q(\boldsymbol{\alpha})}\right)$. It factorizes as $q(\boldsymbol{\gamma}) = \prod_{p=0}^{L-1} q(\gamma[p])$ with its factors being generalized inverse Gaussian pdfs [45]. The first inverse moments of its entries have a closed form expression:

$$\left\langle\gamma[p]^{-1}\right\rangle_{q(\boldsymbol{\gamma})} = \left(\frac{\left\langle|\alpha[p]|^2\right\rangle_{q(\boldsymbol{\alpha})}}{\eta}\right)^{-\frac{1}{2}} \frac{K_{\epsilon-2}\left(2\sqrt{\eta\left\langle|\alpha[p]|^2\right\rangle_{q(\boldsymbol{\alpha})}}\right)}{K_{\epsilon-1}\left(2\sqrt{\eta\left\langle|\alpha[p]|^2\right\rangle_{q(\boldsymbol{\alpha})}}\right)}, \quad p\in[0:L-1], \tag{A.18}$$

where $K_v(\cdot)$ is the modified Bessel function of the second kind and order $v\in\mathbb{R}$.

The belief of the channel weight vector reads $q(\boldsymbol{\alpha}) \propto m^{\mathrm{MF}}_{f_{\mathcal{H}}\to\boldsymbol{\alpha}}(\boldsymbol{\alpha})\prod_{n=0}^{B-1} m^{\mathrm{MF}}_{f_{\mathcal{O}_n}\to\boldsymbol{\alpha}}(\boldsymbol{\alpha})$ where the two factors are $m^{\mathrm{MF}}_{f_{\mathcal{O}_n}\to\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \propto \mathrm{CN}\left(\boldsymbol{\alpha};\; \langle\lambda\rangle_{q(\lambda)}\,\boldsymbol{\Sigma'}_{\boldsymbol{\alpha}}\,\langle\mathbf{A}_n\rangle^{\mathrm{H}}_{q(\mathbf{x}_n),q(\mathbf{x}_{n-1})}\,\mathbf{y}_n,\; \boldsymbol{\Sigma'}_{\boldsymbol{\alpha}}\right)$ with $\boldsymbol{\Sigma'}_{\boldsymbol{\alpha}} = \left(\langle\lambda\rangle_{q(\lambda)}\left\langle\mathbf{A}_n^{\mathrm{H}}\mathbf{A}_n\right\rangle_{q(\mathbf{x}_n),q(\mathbf{x}_{n-1})}\right)^{-1}$, and $m^{\mathrm{MF}}_{f_{\mathcal{H}}\to\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \propto \mathrm{CN}(\boldsymbol{\alpha};\mathbf{0},\langle\boldsymbol{\Gamma}^{-1}\rangle^{-1}_{q(\boldsymbol{\gamma})})$. The product of Gaussian pdfs can be recast as a Gaussian pdf, i.e. $q(\boldsymbol{\alpha}) \propto \mathrm{CN}(\boldsymbol{\alpha};\boldsymbol{\mu}_{\boldsymbol{\alpha}},\boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$ with

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \left(\langle\lambda\rangle_{q(\lambda)}\sum_{n=0}^{B-1}\left\langle\mathbf{A}_n^{\mathrm{H}}\mathbf{A}_n\right\rangle_{q(\mathbf{x}_n),q(\mathbf{x}_{n-1})} + \left\langle\boldsymbol{\Gamma}^{-1}\right\rangle_{q(\boldsymbol{\gamma})}\right)^{-1} \tag{A.19a}$$

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \langle\lambda\rangle_{q(\lambda)}\,\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\sum_{n=0}^{B-1}\langle\mathbf{A}_n\rangle^{\mathrm{H}}_{q(\mathbf{x}_n),q(\mathbf{x}_{n-1})}\,\mathbf{y}_n. \tag{A.19b}$$

**Separate update of the channel weights**

Since the updates (A.19) are computationally expensive, we propose next a solution that avoids large matrix inversions and the evaluation of complicated updates. To that end, we utilize two approximations: (i) we employ the naive MF approximation [39] by assuming the belief of $\boldsymbol{\alpha}$ to factorize as $q(\boldsymbol{\alpha}) = \prod_{i=0}^{L-1} q(\alpha[i])$; (ii) we restrict the belief of $\boldsymbol{\gamma}$ to a Dirac delta function at the mode $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}[0],...,\hat{\gamma}[L-1]]$ of $q(\boldsymbol{\gamma})$ obtained in the paragraph preceding (A.18): $\hat{q}(\boldsymbol{\gamma}) = \prod_{i=0}^{L-1}\hat{q}(\gamma[i]) = \prod_{i=0}^{L-1}\delta(\gamma[i]-\hat{\gamma}[i]), \hat{\gamma}[i] = \mathrm{argmax}_{\gamma[i]}q(\gamma[i])$.

As a consequence of (i), each of the scalar variables $\alpha[i]$ and $\gamma[i]$ is represented with its own variable node in the corresponding factor graph, as depicted in Fig. A.2. Here we define $f_{\mathcal{H}_i}(\alpha[i],\gamma[i]) = p(\alpha[i]|\gamma[i])$ and $f_{\mathcal{V}_i}(\gamma[i]) = p(\gamma[i])$, with $p(\alpha[i]|\gamma[i]) = \mathrm{CN}(\alpha[i];0,\gamma[i])$ and $p(\gamma[i]) = \mathrm{Ga}(\gamma[i];\epsilon,\eta)$.

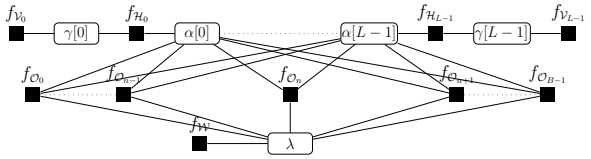


**Fig. A.2:** MF subgraph corresponding to the fully factorized belief of the channel weights.

Following (ii), we substitute the first inverse moment updates for the entries in $\boldsymbol{\gamma}$ in (A.18) by updates more easy to evaluate, i.e the inverse of

$$\hat{\gamma}[i] = \frac{\epsilon - 2 + \sqrt{(\epsilon-2)^2 + 4\eta \left\langle |\alpha[i]|^2 \right\rangle_{q(\alpha[i])}}}{2\eta}, \; i \in [0 : L-1]. \tag{A.20}$$

Consequently, the belief of each channel weight $\alpha[i]$ reads $q(\alpha[i]) \propto m_{f_{\mathcal{H}_i} \to \alpha[i]} \prod_{n=0}^{B-1} m_{f_{\mathcal{O}_n} \to \alpha[i]} \propto \mathrm{CN}(\alpha[i]; \mu_\alpha[i], \sigma_\alpha^2[i])$ with

$$\sigma_\alpha^2[i] = \left( \langle \lambda \rangle_{q(\lambda)} \, \varsigma[i] + \hat{\gamma}[i]^{-1} \right)^{-1} \tag{A.21a}$$

$$\mu_\alpha[i] = \langle \lambda \rangle_{q(\lambda)} \, \sigma_\alpha^2[i] \theta[i] \tag{A.21b}$$

where $\varsigma[i] = \sum_{n=0}^{B-1} \left\langle \| \, \mathbf{A}_n[\cdot, i] \, \|_2^2 \right\rangle_{q(\mathbf{x}_n), q(\mathbf{x}_{n-1})}$ and $\theta[i] = \sum_{n=0}^{B-1} \big( \langle \mathbf{A}_n^*[i, \cdot] \rangle_{q(\mathbf{x}_n), q(\mathbf{x}_{n-1})} \, \mathbf{y}_n - \sum_{p=0, p \neq i}^{L-1} \langle \mathbf{A}_n^*[i, \cdot] \mathbf{A}_n[\cdot, p] \rangle_{q(\mathbf{x}_n), q(\mathbf{x}_{n-1})} \cdot \langle \alpha[p] \rangle_{(q(\alpha[p]))} \big)$.

The separate weights estimation scheme updates consecutively the belief of each channel weight $\alpha[i]$ by iteratively computing $n_{\alpha[i] \to f_{\mathcal{H}_i}}$, $m^{\mathrm{MF}}_{f_{\mathcal{H}_i} \to \gamma[i]}$, $n_{\gamma[i] \to f_{\mathcal{V}_i}}$, $m^{\mathrm{MF}}_{f_{\mathcal{V}_i} \to \gamma[i]}$, $n_{\gamma[i] \to f_{\mathcal{H}_i}}$ and $m^{\mathrm{MF}}_{f_{\mathcal{H}_i} \to \alpha[i]}$ until convergence, while keeping the beliefs of $\alpha[j], j \neq i$ fixed. Since many iterations are needed before the convergence of each belief $q(\alpha[i])$ is achieved, we propose a recursive update of $\gamma[i]$ and consequently of $\alpha[i]$, inspired by [46]. The above iterative computation of messages can be expressed as a recursive procedure to update the estimates of $\boldsymbol{\gamma}$, which can be written in the form $\hat{\gamma}[i]^{new} = g_i(\hat{\gamma}[i]^{old})$. The function $g_i$ characterizing the recursion can be obtained by inserting (A.21b) and (A.21a) in (A.20), which yields

$$g_i(\hat{\gamma}[i]) = \frac{\epsilon - 2}{2\eta} + \sqrt{\left( \frac{\epsilon - 2}{2\eta} \right)^2 + \frac{\hat{\gamma}[i] \left( (c_i + q_i) \hat{\gamma}[i] + c_i^2 \right)}{\eta (\hat{\gamma}[i] + c_i)^2}} \tag{A.22}$$

where $c_i^{-1} = \lambda \varsigma[i]$ and $q_i = \varsigma[i]^{-2} |\theta[i]|^2$. If the iterative computation of messages described above converges, it does it to a fixed point of $g_i$. We need thus the solutions of $\hat{\gamma}^*[i] = g_i(\hat{\gamma}^*[i])$, which corresponds to solving the cubic

equation

$$\begin{aligned}&\hat{\gamma}^*[i]\left(\eta\hat{\gamma}^*[i]^3+\left[2\eta c_i-(\epsilon-2)\right]\hat{\gamma}^*[i]^2+\right.\\&\left.\left[\eta c_i^2-c_i(2\epsilon-3)-q_i\right]\hat{\gamma}^*[i]-(\epsilon-1)c_i^2\right)=0.\end{aligned}\tag{A.23}$$

For the discussion on finding the solution of (A.23), see Appendix A. Once the fixed point $\hat{\gamma}^*[i]$ has been obtained, the belief of $\alpha[i]$ is updated using (A.21b) and (A.21a).

## 3.3 Description of the receiver algorithm

It can be seen from the updates presented in Section 3.2 that computing the beliefs of the current $(n)$ data symbols $x_n[d], d \in \mathcal{D}$ implies computing messages from nodes corresponding to the next $(n+1)$, and previous $(n-1)$ signaling intervals in the transmission block. Consequently, there are multiple scheduling options i.e. selections of the order in which the different messages, and hence the updates of the beliefs, are performed.

In one instance, one could choose to wait until all $\mathbf{y}_n, n \in [0:B-1]$ signals in the block have been received and estimate all $\mathbf{u}_n, n \in [0:B-1]$ using a *forward-backward* scheduling, with messages propagating in a first stage from left to right in the factor graph in Fig. A.1 and in a second stage from right to left. This scheduling, however, is computationally and memory expensive, in addition to implying some latency. Its implementation is therefore impractical in real-time receivers. As an alternative which overcomes the aforementioned disadvantages of the forward-backward algorithm, we propose a scheduling that performs an instance of the *forward* stage only. An additional virtue of the later scheduling is that the estimate of $\mathbf{u}_n$ can be obtained once the observation $\mathbf{y}_n$ is available.

Algorithm 1 implements the *forward* processing and thus works sequentially, i.e. at each signaling interval $n$, the algorithm makes use of the observations $\boldsymbol{y}_k, k \leq n$ and the soft estimates of $\mathbf{x}_{k'}, k' \leq n-1$. Then, it updates the estimates of the channel weights, noise precision and current symbol vector $\mathbf{x}_n$ and consequently $\mathbf{u}_n$. This scheduling results from setting all future messages, i.e. messages that require access to variables from future signaling intervals, to constants. Specifically, in the computation of $q(x_n[d_k])$ in (A.11), $q(\lambda)$ before (A.17), and $q(\boldsymbol{\alpha})$ before (A.19), we set: $m^{\mathrm{MF}}_{f_{\mathcal{O}_{n+1}}\to x_n[d_k]}(x_n[d_k])=1$, $\prod_{k=n+1}^{B-1} m^{\mathrm{MF}}_{f_{\mathcal{O}_k}\to\lambda}(\lambda)=1$, and $\prod_{k=n+1}^{B-1} m^{\mathrm{MF}}_{f_{\mathcal{O}_k}\to\boldsymbol{\alpha}}(\boldsymbol{\alpha})=1$, respectively. The algorithm iterates between data and channel updates until convergence is reached, or the number of performed iterations exceeds a maximum value. We define two types of iterations: (1) an outer iteration – which is completed when an estimate of the current data symbols $\mathbf{x}_n$ is obtained, and (2) an inner iteration – which is completed when an estimate of the channel weights $\boldsymbol{\alpha}$ is obtained. We denote the maximum number of outer and inner iterations as $M_o$ and $M_i$

**Algorithm 1** Proposed receiver algorithm

set $\Delta\tau$, $L$ and $\langle x_n[d_k]\rangle_{q(x_n[d_k])} = 0$, $\left\langle |x_n[d_k]|^2\right\rangle_{q(x_n[d_k])} = 0$, $k \in [0 : N_D - 1]$, $n = [0 : B-1]$
for current symbol $i \in [0 : B-1]$ do
initialize $\langle\lambda\rangle_{q(\lambda)}$, $\left\langle\gamma^{-1}\right\rangle_{q(\boldsymbol{\gamma})}$, $\langle\boldsymbol{\alpha}\rangle_{q(\boldsymbol{\alpha})}$, $\left\langle\boldsymbol{\alpha}\boldsymbol{\alpha}^{\mathrm{H}}\right\rangle_{q(\boldsymbol{\alpha})}$ and the bounds for outer ($M_o[i]$) and inner ($M_i[i]$) iterations
**for** $IT = 1 : M_o[i]$ **do**
  **for** $it = 1 : M_i[i]$ **do**
    compute $\langle\lambda\rangle_{q(\lambda)}$ using (A.17)
    **if** (joint $\boldsymbol{\alpha}$ update - scheme 3.2) **then**
      compute $\langle\boldsymbol{\gamma}\rangle_{q(\boldsymbol{\gamma})}$ with (A.18) and $\boldsymbol{\mu}_\alpha$, $\boldsymbol{\Sigma}_\alpha$ with (A.19)
    **else if** (separate $\boldsymbol{\alpha}$ update - scheme 3.2) **then**
      **for** $p \in [0 : L-1]$ **do**
        compute $\langle\gamma[p]\rangle_{q(\gamma[p])}$ with (A.23) and $\mu_\alpha[p]$, $\sigma^2_\alpha[p]$, with (A.21)
      **end for**
    **end if**
  **end for**
  compute $\langle x_i[d_k]\rangle_{q(x_i[d_k])}$, $\left\langle |x_i[d_k]|^2\right\rangle_{q(x_i[d_k])}$, with (A.14), (A.16) and $\langle\mathbf{A}_i\rangle_{q(\mathbf{x}_i),q(\mathbf{x}_{i-1})}$ with (A.7) and $\left\langle\mathbf{A}_i^{\mathrm{H}}\right\rangle_{q(\mathbf{x}_i),q(\mathbf{x}_{i-1})}\mathbf{y}_i$ and $\left\langle\mathbf{A}_i^{\mathrm{H}}\mathbf{A}_i\right\rangle_{q(\mathbf{x}_i),q(\mathbf{x}_{i-1})}$
**end for**

respectively. Note that the higher the index of the signaling interval, the more observation vectors are available for estimating the channel weights and the noise precision. Therefore computation of the channel updates at a given signaling interval may need fewer iterations than in the previous intervals. To account for that, one may set a number of maximum iterations that decreases with the signaling index.

## 3.4 Computational complexity

We discuss next the main sources of complexity of Algorithm 1, together with potential complexity reductions alternatives.

**Updating the channel weights** The complexity of computing $\boldsymbol{\mu}_\alpha$ and $\boldsymbol{\Sigma}_\alpha$ using the joint update method from (A.19) are $\mathcal{O}(L^2)$ and $\mathcal{O}(L^3)$. The complexity of the matrix inversion involved in the computation of $\boldsymbol{\Sigma}_\alpha$ can be reduced by invoking Woodbury matrix identity when $M < L$: after re-writing the matrix $\langle\lambda\rangle_{q(\lambda)}\sum_{n=0}^{B-1}\left\langle\mathbf{A}_n^{\mathrm{H}}\mathbf{A}_n\right\rangle_{q(\mathbf{x}_n),q(\mathbf{x}_{n-1})} = \boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{\Sigma}_n\boldsymbol{\Phi}$, the covariance matrix becomes[6] $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \boldsymbol{\Delta}^{-1} - \boldsymbol{\Delta}^{-1}\boldsymbol{\Phi}^{\mathrm{H}}\mathbf{K}_n^{-1}\boldsymbol{\Phi}\boldsymbol{\Delta}^{-1}$, where $\mathbf{K}_n = {\boldsymbol{\Sigma}_n}^{-1} + \boldsymbol{\Phi}\boldsymbol{\Delta}^{-1}\boldsymbol{\Phi}^{\mathrm{H}}$ is inverted at a cost $\mathcal{O}(M^3)$. Using the separate update method from (A.21) we compute the entries of the vectors $\boldsymbol{\mu}_\alpha$ and $\boldsymbol{\sigma}^2_\alpha$, of complexity $\mathcal{O}(L)$ each.

[6] We define $\boldsymbol{\Sigma}_n = \langle\lambda\rangle_{q(\lambda)}\sum_{n=0}^{B-1}\left\langle\mathbf{V}_n^{\mathrm{H}}\mathbf{V}_n\right\rangle_{q(\mathbf{x}_n),q(\mathbf{x}_{n-1})} \in \mathbb{C}^{M\times M}$ and the diagonal matrix $\boldsymbol{\Delta} = \left\langle\boldsymbol{\Gamma}^{-1}\right\rangle_{q(\boldsymbol{\gamma})}$.

Alternatively, by choosing a different factorization of the beliefs of the channel weights, similar to that proposed in [47], e.g. $q(\boldsymbol{\alpha}) = \prod_{g=1}^{g=L/z} q(\boldsymbol{\alpha}_g)$, where the vector $\boldsymbol{\alpha}_g$ has $z < L$ entries, we can reduce the problem to estimating $L/z$ groups of channel weights, and thus computing $L/z$ variance matrices $\boldsymbol{\Sigma}_{\alpha_g}$ of complexity $\mathcal{O}(z^3)$. Using the separate update method in (A.21) is equivalent to setting $z = 1$.

**Learning the dictionary** This step consists of recomputing the products $\left\langle \mathbf{A}_i^{\mathrm{H}} \right\rangle_{q(\mathbf{x}_i),q(\mathbf{x}_{i-1})} \mathbf{y}_i$ and $\left\langle \mathbf{A}_i^{\mathrm{H}} \mathbf{A}_i \right\rangle_{q(\mathbf{x}_i),q(\mathbf{x}_{i-1})}$ at the costs $\mathcal{O}(NL)$ and $\mathcal{O}(NL^2)$ respectively. The products are used in (A.19) or (A.21), and (A.17) to refine the channel and noise precision estimates. The two updates can be performed either (i) after each or (ii) all data symbol estimate(s) has/have been obtained. To avoid the complexity cost incurred in case (i), i.e. performing channel estimation $N_D$ times with complexity described in the above paragraph, Algorithm 1 implements variant (ii).

# 4 Performance evaluation

We consider a SISO CP-OFDM system with the parameters given in Table A.2. The CP duration, subcarrier spacing and pilot density are chosen following [6]. We use a stochastic channel model with a long CIR inspired by the

**Table A.2:** OFDM system parameters.

| N | Pilot density & pattern | Modulation | Subcarrier space [kHz] | Channel coding | $T_{CP}$ [$\mu s$] | Block [no. symbols] |
|---|---|---|---|---|---|---|
| 64 | 17% equally spaced | QPSK | 15 | 1/3 $(133, 171, 165)_8$ | 4.69 | 4 |

measurement campaigns reported in [7, 9]. Following the model proposed in [8] the CIR consists of two clusters. Cluster (1) contains the multipath components with delays smaller than the CP duration, while cluster (2) comprises the components with delays beyond the CP length. Cluster (1) is the result of the propagation medium between the serving base station and the mobile terminal, whereas cluster (2) exists due to the reflections from distant tall scatterers. The CIR over the block duration can be written as $\bar{g}(\tau) = \sum_{l=0}^{\bar{L}_{(1)}-1} \bar{\alpha}_{(1)}[l]\delta(\tau - \bar{\tau}_{(1)}[l]) + \sum_{l=0}^{\bar{L}_{(2)}-1} \bar{\alpha}_{(2)}[l]\delta(\tau - \bar{\tau}_{(2)}[l])$, where $\bar{\boldsymbol{\alpha}}_{(k)}$ and $\bar{\boldsymbol{\tau}}_{(k)}$ contain the weights and delays of the $\bar{L}_{(k)}$ multipath components in cluster $(k), k \in [1:2]$. We define the power-delay profile (PDP): $P(\tau) = Q\exp(-\frac{\tau}{v})I_{[0,T_{CP}]}(\tau) + aQ\exp(\frac{T_{CP}-\tau}{v})I_{(T_{CP},\bar{\tau}_M]}(\tau)$ where $v = 1\mu s$, $a$ is the power attenuation factor of the cluster (2), $Q$ is a scaling parameter, and $\bar{\tau}_M$ is the channel MED. The joint pdf of the unknown parameters is $p\left(\left[\bar{\boldsymbol{\alpha}}_{(k)}, \bar{\boldsymbol{\tau}}_{(k)}, \bar{L}_{(k)}; k \in [1:2]\right]\right) =$

**Table A.3:** Receivers used for benchmarking. All receivers use BCJR decoding. Channel estimation in a given signaling interval, say $n$, is performed by using all observation vectors up to $n$: $\mathbf{y}_k, k \leq n$.

| Receiver | Channel Estimation | Equalization |
|---|---|---|
| PSER | pilot-based channel estimation using SBI | one-tap equalizer (ISI/ICI are ignored) |
| SIC | perfect channel state information (CSI) | successive interference cancellation [2] |
| PSE-SIC | pilot-based channel estimation using SBI | successive interference cancellation [2] |
| GAR | perfect CSI | LMMSE equalization with perfect ISI & ICI cancellation and knowledge of noise precision |

$\prod_{k=1}^{2} p(\bar{\boldsymbol{\alpha}}_{(k)}|\bar{\boldsymbol{\tau}}_{(k)})p(\bar{\boldsymbol{\tau}}_{(k)}|\bar{L}_{(k)})p(\bar{L}_{(k)})$, where $p(\bar{\boldsymbol{\tau}}_{(k)}|\bar{L}_{(k)}) = \prod_{l=0}^{\bar{L}_{(k)}-1} p(\bar{\tau}_{(k)}[l])$. The number of components $\bar{L}_{(k)}$ is Poisson distributed $\bar{L}_{(k)} \sim \text{Poisson}(\mu_{\bar{L}_k})$. The delays are independent and uniformly distributed, i.e. $\bar{\tau}_{(1)}[l] \sim \text{Unif}(0, T_{CP})$, $\bar{\tau}_{(2)}[l] \sim \text{Unif}(T_{CP}, \bar{\tau}_M)$. The conditional pdf of $\bar{\boldsymbol{\alpha}}_{(k)}$ reads $p(\bar{\boldsymbol{\alpha}}_{(k)}|\bar{\boldsymbol{\tau}}_{(k)}) = \prod_{l=0}^{\bar{L}_{(k)}-1} \text{CN}\left(\bar{\alpha}_{(k)}[l]|0, P(\bar{\tau}_{(k)}[l])\right)$. We assume $\bar{\tau}_{(1)} = 0$ which corresponds to perfect time synchronization of the receiver. Unless stated otherwise, CIR realizations are generated using the above model with $\bar{\tau}_M = 10\mu s$, $\mu_{\bar{L}_1} = \mu_{\bar{L}_2} = 5$, $a = 1$.

## 4.1 Tested receivers

We evaluate the receiver design proposed in Section 3 with the two different configurations. One receiver, which we abbreviate as MPR-J, implements Algorithm 1 with joint channel weight estimation described in Section 3.2. The second receiver, abbreviated as MPR-S, implements Algorithm 1 with separate (disjoint) channel weights estimator from Section 3.2. We set $\epsilon = 0.5$ and $\eta = 1$ as in [48], and initialize $\langle \boldsymbol{\gamma}^{-1} \rangle_{q(\boldsymbol{\gamma})} = 1$ and $\langle \lambda \rangle_{q(\lambda)} = 1$. We benchmark the bit-error rate (BER) performance of the two proposed receivers against that of the receivers listed in Table A.3. GAR and PSER represent respectively the lower and upper performance bounds. SIC and PSE-SIC have been chosen to highlight how the errors stemming from imperfect channel estimation and detection, impact the overall BER: since SIC works with perfect CSI, the detection block is the sole source of error; because PSE-SIC treats ISI and ICI as AWGN when estimating the channel weights, we expect its performance to be further impaired by channel estimation errors. The purpose of evaluating SIC and PSE-SIC is to illustrate the necessity of a joint design of channel estimator and equalizer, since an equalizer that performs close to optimally with perfect CSI may, however, perform very poorly when inaccurate channel estimates are

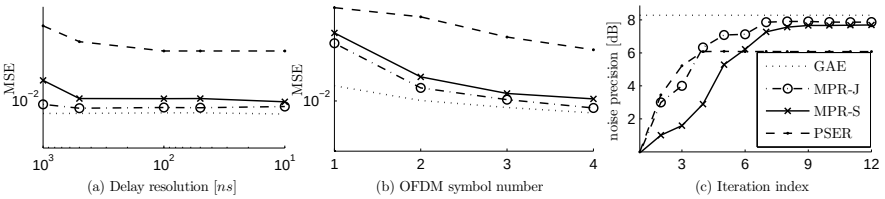


**Fig. A.3:** (a) MSE at the last symbol in the block versus the delay resolution, (b) MSE versus the OFDM symbol number and (c) estimated noise precision versus the number of total iterations at 8 dB SNR. The legend applies to all figures.

used instead. To test the performance of the channel estimators implemented in MPR-J and MPR-S, their MSE are also evaluated and compared to the MSE of the estimator in PSER, as well as of a genie-aided MMSE channel estimator (GAE) that knows all the symbols in the block, the delays of all multipath components and the noise variance.

Three design parameters need to be set for MPR-J, MPR-S and likewise for PSER: (i) the delay resolution, (ii) the MED assumed by the CIR estimator, (iii) and the number of iterations of the algorithm. Based on a survey of the literature on channel measurements [7, 9, 10], we assume $\max(\boldsymbol{\tau}_s) = 10\mu s$. To investigate the influence of (i) and (iii), we assess the performance of the proposed receivers in the system specified in Table A.2. With this setup, we vary the delay resolution of the receivers and observe the MSE performance. In Fig. A.3.a we observe that the MSE of all channel estimators reaches its minimum value for a delay resolution $\Delta\tau \leq 100ns$. Fig. A.3.b indicates that all receivers benefit from the increased number of observations, however, MPR-J and MPR-S exhibit the largest MSE improvement due to the data-aided nature of their CIR estimators. To set the number of iterations, we monitor the noise precision estimate which is a good convergence indicator since it measures the distance between the observed signal and the signal reconstructed using the channel and symbol estimates. Fig. A.3.c depicts the estimated noise precision over the total number of iterations[7].

## 4.2 Numerical study

### Varying the power attenuation factor of cluster (2)

Fig. A.4 depicts the BER of the proposed receivers and the benchmarkers for three different values of the power attenuation factor: $a = 0.1, a = 0.5$, inspired by the HT and BU models [8], and $a = 1$ which emulates very harsh

[7]By total number of iterations we mean the overall number of channel and noise precision updates since the initialization of the algorithm. In the simulations we set the total number of iterations for MPR-S, MPR-J and PSER to 12.

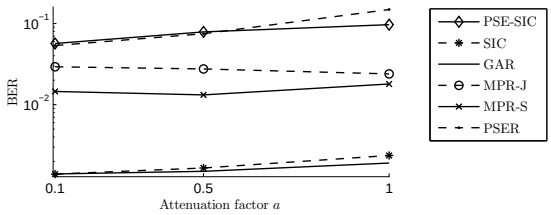


**Fig. A.4:** BER at SNR= 4 dB versus attenuation factor $a$.

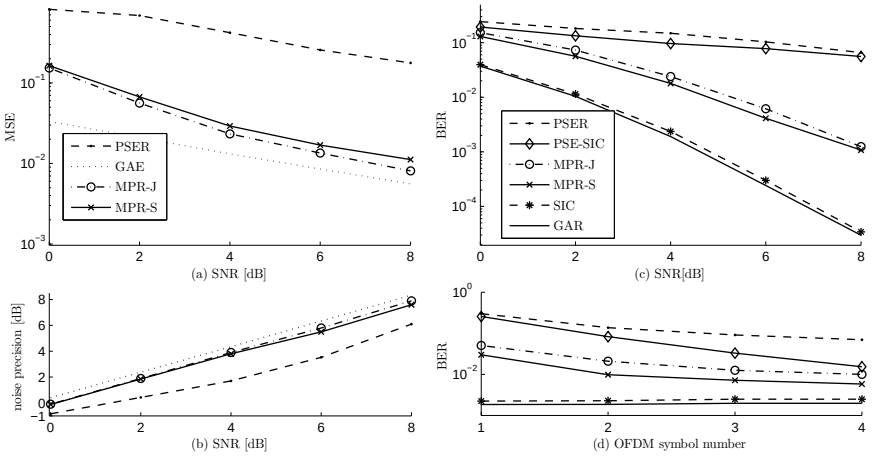


**Fig. A.5:** (a) - (b) MSE and estimated noise precision versus SNR (legend applies to both figures). Fig. (c) - (d) BER versus SNR, and BER at SNR= 4 dB versus the OFDM symbol number (legend applies to both figures). The attenuation factor is $a = 1$.

propagation conditions similar to those reported in [9]. MPR-J and MPR-S exhibit a robust behaviour w.r.t. $a$, the latter exhibiting only a slight BER increase at $a = 1$. By contrast, PSE-SIC and PSER are more sensitive to ISI and ICI, their BER displaying a steady degradation.

Fig. A.5.a and Fig. A.6.a show that the channel estimator of MPR-J exhibits a marginally better MSE performance than MPR-S at high SNR, which could be attributed to the joint estimation of the channel weights. Unlike PSER, both MPR-J and MPR-S correctly estimate the noise precision, see Fig. A.5.b. The erroneous estimate returned by PSER results from both the poor quality of the channel estimate and the lack of an interference cancellation scheme. As depicted in Fig. A.5.c and Fig. A.6.b both MPR-J and MPR-S clearly outperform PSER and PSE-SIC in terms of BER. Due to the poor accuracy of the pilot-based channel estimates, PSE-SIC does not benefit from sequentially canceling the interference and shows a BER similar to that of PSER. Lastly Fig. A.5.d shows that increasing the observation size

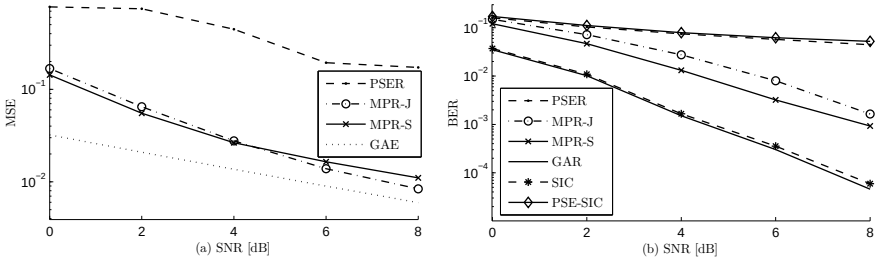


**Fig. A.6:** (a) - (b) MSE and BER versus SNR. The attenuation factor is $a = 0.5$.

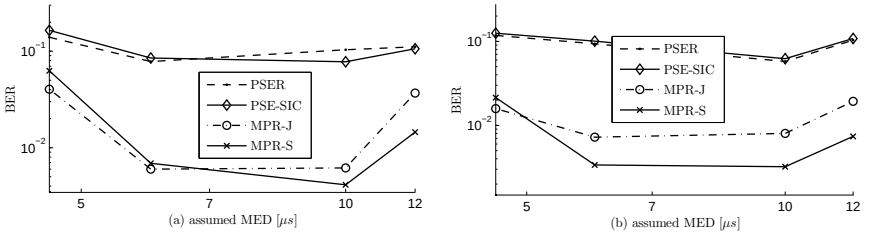


**Fig. A.7:** BER at SNR= 6 dB versus the assumed MED for attenuation factors: (a) $a = 1$ and (b) $a = 0.5$.

leads to an increase in the number of pilot symbols which determines the BER improvement of PSER and PSE-SIC.

### Varying the assumed MED

In this study we vary $\max(\boldsymbol{\tau}_s)$ in the interval $[4.6,\ 12]\mu s$. The BER performance is shown in Fig. A.7.a and Fig. A.7.b for two values of the attenuation factor. When $\max(\boldsymbol{\tau}_s) = 4.6\mu s$, the receivers cannot detect the multipath components in cluster (2). The slight performance improvement that MPR-J and MPR-S exhibit relative to PSER and PSE-SIC is due to the data-aided nature of the channel estimation. As the assumed MED approaches the true MED, $\bar{\tau}_M = 10\mu s$, the proposed receivers start accounting for ISI and ICI and show a rapid improvement of their BER. Even though the delay range in which the channel estimators look for multipath components is smaller than the true one, MPR-J and MPR-S are able to find an equivalent channel representation which cancels the interference and enables a reliable decoding of the data symbols. Selecting $\max(\boldsymbol{\tau}_s) > \bar{\tau}_M$ leads to performance degradation, less significant as $a$ decreases. Concluding, MPR-J and MPR-S are robust to underestimation, as long as the assumed MED is larger than the CP length.

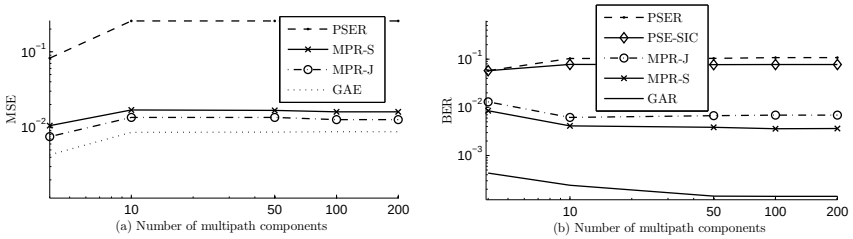


**Fig. A.8:** (a)-(b) MSE and BER performance at SNR= 6 dB versus number of multipath components.

### Increasing the number of multipath components

We generate channel realizations with increasing number of multipath components in each cluster. Fig. A.8 shows that, as the channel becomes less sparse – i.e. the number of multipath components increases – the MSE of the channel estimates for all receivers slightly degrades, as does the MSE bound provided by GAE. Hence, even though the assumption of a sparse channel response is no longer fulfilled, the estimators in MPR-J and MPR-S still yield an accuracy very close to that of an ideal estimator. The effect is opposite in terms of BER performance: the increased degrees of freedom, due to the growing number of multipath components, enable the receivers' decoders to resolve the data symbols with increased accuracy. We note nonetheless that the proposed receivers exhibit an overall robust behavior w.r.t. the level of sparsity of the channel. The reason for this is that, even when the true response holds tens of multipaths components, the channel estimator can still find a sparse representation that closely approximates the true channel response with only a few components.

# 5 Conclusion

To overcome the performance degradation experienced by the traditional OFDM receiver when operating in insufficient CP conditions, we propose an iterative algorithm capable of reconstructing and canceling the self-interference. Numerical evaluations show that the receivers implementing the proposed algorithm perform close to a genie-aided receiver, even under particularly harsh propagation conditions. By contrast, due to the separate processing of channel estimation and data detection, the benchmark receivers suffer noticeable degradations. This finding indicates that a joint design of all receiver blocks like the one we propose, is necessary for attaining high BER performance.

Our design is versatile in several respects: the variety of schedules that can be selected for updating the different messages and beliefs, the number of

OFDM symbols that can be processed jointly in a block, and the implementation of various channel estimation schemes, e.g. group or separate estimation of the channel weights. This diversity of options allows for trading off computational complexity for performance – and vice-versa – in the algorithm design to better adjust to the practical constraints imposed by the considered system. Furthermore, the proposed scheme can be turned on and off depending on the channel conditions. It suffices that the receiver is equipped with an estimator of the maximum excess delay at the cost of a slight complexity increase compared to classical receivers.

To conclude, we stress that our solution offers new perspectives in terms of the waveform design for future cellular systems like 5G. Given the scarcity of the spectrum, the 5G systems need to be optimized for both spectral efficiency and high data rate. The former requirement implies relaxing the strict subcarrier orthogonality, making the later requirement even harder to achieve without equipping the receiver with self-interference cancellation capability. This is precisely what our receiver does. In a system with reduced CP length and receivers equipped with this capability, all users served by the same base station experience a similar quality of service/transmission performance regardless of whether the maximum excess delay of their channel exceeds the CP length or not.

# A Computing the fixed points of (A.22)

We re-write[8] (A.23) as $\gamma f(\gamma) = 0$ where $f(\gamma) = \eta\gamma^3 + [2\eta c - (\epsilon - 2)]\,\gamma^2 + \left[\eta c^2 - c(2\epsilon - 3) - q\right]\gamma - (\epsilon - 1)c^2$. Clearly, $\gamma = 0$ is always a fixed point of (27), so we focus the following discussion on computing the fixed points corresponding to the roots of $f(\gamma)$. In this work, we are interested in the case $0 \leq \epsilon \leq 1$ and $\eta > 0$ that we analyze below. For a more elaborate study and other setting of $\epsilon$ and $\eta$, we refer to [48]. Since $\lim_{\gamma\to-\infty} f(\gamma) = -\infty$, and $f(0) \geq 0$, $f(\gamma)$ has at least one negative root. To find the two remaining roots, we make use of the fact that $\lim_{\gamma\to\infty} f(\gamma) = \infty$ and analyze the stationary points of $f(\gamma)$, denoted by $s^{(1)}$ and $s^{(2)}$. If at least one of the stationary points is complex, then $f(\gamma)$ has at least one complex root (and consequently its conjugate), and therefore, no positive root exists [49, Gauss-Lucas Thm]. If both stationary points $s^{(1)}, s^{(2)}$ are real and $s^{(2)} < 0$ (where, without loss of generality we assume $s^{(1)} \leq s^{(2)}$), then $f(\gamma)$ has no positive root. If $s^{(2)} > 0$, $f(\gamma)$ can have one or two positive roots.

To summarize, $f(\gamma)$ has either no positive root (when $s^{(2)} \geq 0$ and $f(s^{(2)}) > 0$ or $s^{(2)} < 0$) or two positive roots[9] (when $s^{(1)} \leq s^{(2)}, s^{(2)} \geq 0$ and $f(s^{(2)}) \leq 0$).

[8] For the ease of notation, we discard the index $i$ in both the function argument and its coefficients and we replace $\hat{\gamma}^*[i]$ by $\gamma$.

[9] If $f(s^{(2)}) = 0$, then $s^{(2)}$ is a double root.

If $f(\gamma)$ has no positive root, then we select $\gamma^* = 0$ as the fixed point.[10] When $f(\gamma)$ has two positive roots, we select the fixed point corresponding to the largest root of $f(\gamma)$.

# B Derivation of the expectations used in section 3.2

The matrices $\langle \mathbf{M} \rangle_{q(\boldsymbol{\alpha})}, \langle \mathbf{N} \rangle_{q(\boldsymbol{\alpha})}$, and $\langle \mathbf{A}_n \rangle_{q(\mathbf{x}_{n-1})q(\mathbf{x}_n)}$ are obtained by replacing $\boldsymbol{\alpha}$, $\mathbf{x}_{n-1}, \mathbf{x}_n$ in the definitions after (A.8) by their respective expected values over their belief as obtained in (A.19) - or (A.21) - and (A.11). To update $\left\langle \mathbf{A}_n^{\mathrm{H}} \mathbf{A}_n \right\rangle_{q(\mathbf{x}_{n-1})q(\mathbf{x}_n)}$, $\left\langle \mathbf{M}^{\mathrm{H}} \mathbf{M} \right\rangle_{q(\boldsymbol{\alpha})}$ and $\left\langle \mathbf{N}^{\mathrm{H}} \mathbf{N} \right\rangle_{q(\boldsymbol{\alpha})}$, we compute the diagonal matrix $\mathbf{S}_{2H} = \left\langle \mathbf{H}^{\mathrm{H}} \mathbf{H} \right\rangle_{q(\boldsymbol{\alpha})}$, the matrices $\mathbf{S}_{2C} = \left\langle \mathbf{C}^{\mathrm{H}} \mathbf{C} \right\rangle_{q(\boldsymbol{\alpha})}$, $\mathbf{S}_{HC} = \left\langle \mathbf{C}^{\mathrm{H}} \mathbf{F}^{\mathrm{H}} \mathbf{H} \right\rangle_{q(\boldsymbol{\alpha})}$, $\mathbf{S}_{2S} = \left\langle \mathbf{S}^{\mathrm{H}} \mathbf{S} \right\rangle_{q(\boldsymbol{\alpha})}$ from Table A.4 and $\left\langle \boldsymbol{\Xi}_n^{\mathrm{H}} \boldsymbol{\Xi}_n \right\rangle_{q(\mathbf{x}_{n-1})q(\mathbf{x}_n)} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{2\mathrm{Y}_n} \end{bmatrix}$, $\mathbf{S}_{2\mathrm{Y}_n} = \left\langle \mathbf{Y}_n^{\mathrm{H}} \mathbf{Y}_n \right\rangle_{q(\mathbf{x}_{n-1})q(\mathbf{x}_n)}$.

**Table A.4:** Matrices definitions

| |
|---|
| $S_{2H}[i,i] = N \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} (F\Phi)[i,l](F\Phi)^*[i,l'] S_\alpha[l,l']$ |
| $S_{2C}[i,v] = \sum_{m=0}^{E-1} \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} \Phi[N+m-i,l]\Phi[N+m-v,l'] S_\alpha[l,l'] \cdot$ |
| $I_{[m+N-E-\mu:N-\mu-1]}(i) I_{[m+N-E-\mu:N-\mu-1]}(v)$ |
| $S_{2S}[i,v] = \sum_{m=0}^{E-1} \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} \Phi[N+\mu+m-i,l]\Phi[N+\mu+m-v,l'] S_\alpha[l,l'] \cdot$ |
| $I_{[m+N-E:N-1]}(i) I_{[m+N-E:N-1]}(v)$ |
| $S_{HC}[i,v] = -\sqrt{N} \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} S_\alpha[l,l'] \sum_{a=0}^{E-1} \Phi[N+\mu+a-i,l'] \sum_{k=0}^{M-1} \Phi[k,l] F^*[v,a] \cdot$ |
| $F[v,k] I_{[a+N-E-\mu:N-\mu-1]}(i)$ |
| $S_{2\mathrm{Y}_n}[a,b] = \sum_{m,i,j=0}^{N-1} \Big( S_{x_{n-1}}[i,j] \varpi^*[j,j] \varpi[i,i] - \varpi^*[j,j] \left\langle x_{n-1}^*[j] \right\rangle_{q(\mathbf{x}_{n-1})} \langle x_n[i] \rangle_{q(\mathbf{x}_n)} -$ |
| $\varpi^*[i,i] \langle x_{n-1}[i] \rangle_{q(\mathbf{x}_{n-1})} \langle x_n^*[i] \rangle_{q(\mathbf{x}_n)} + S_{x_n}[i,j] \Big) (\Lambda^{(m)}[i,a])^* \Lambda^{(m)}[j,b],$ |

We define $\mathbf{S}_{\boldsymbol{\alpha}} = \left\langle \boldsymbol{\alpha} \boldsymbol{\alpha}^{\mathrm{H}} \right\rangle_{q(\boldsymbol{\alpha})}$. The matrix $\mathbf{S}_{x_m} = \left\langle \mathbf{x}_m \mathbf{x}_m^{\mathrm{H}} \right\rangle_{q(\mathbf{x}_m)}$ is obtained from (A.11).

# References

[1] "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception (Release 8)," 3GPP, Tech. Rep., June 2011.

[10] This is equivalent to removing the corresponding column of the dictionary matrix, as the variance of the associated weight is zero.

[2] A. Molisch, M. Toeltsch, and S. Vermani, "Iterative Methods for Cancellation of Intercarrier Interference in OFDM Systems," vol. 56, pp. 2158–2167, 2007.

[3] W. Lee and C. S. Curry, "A performance analysis of OFDM systems in excessively dispersive multipath channels," *Journal of Commun. and Networks*, vol. 8, pp. 323–329, 2006.

[4] J. Montojo and L. Milstein, "Channel estimation for non-ideal OFDM systems," vol. 58, pp. 146–156, 2010.

[5] W. Lee, "On channel estimation for OFDM systems in multipath environments with relatively large delay spread," in *The 57th IEEE Semiannual Vehicular Techn. Conf. VTC '03*, vol. 2, April 2003, pp. 1303–1307.

[6] "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 10)," 3GPP, Tech. Rep., December 2011.

[7] S. Salous and H. Gokalp, "Dual-frequency sounder for UMTS frequency-division duplex channels," vol. 149, pp. 117–122, 2002.

[8] M. Failli, "Digital land moble radio communications," Commission of the European Communities, Tech. Rep., 1988.

[9] T. Rautiainen, J. Juntunen, and K. Kalliola, "Propagation Analysis at 5.3 GHz in Typical and Bad Urban Macrocellular Environments," in *IEEE 65th Vehicular Techn. Conf. VTC '07-Spring.*, April 2007, pp. 501–505.

[10] P. Kyösti, J. Meinilä, L. Hentilä, X. Zhao, T. Jämsä, C. Schneider, M. Narandzic, M. Milojevic, A. Hong, J. Ylitalo, V.-M. Holappa, M. Alatossava, R. Bultitude, Y. deJong, and T. Rautiainen, "IST-4-027756 WINNER II, D1.1.1 v1.1 interim channel models," Information Society Technologies, Tech. Rep., 2006.

[11] H. Steendam and M. Moeneclaey, "Analysis and optimization of the performance of OFDM on frequency-selective time-selective fading channels," vol. 47, pp. 1811–1819, 1999.

[12] F. Schaich, T. Wild, and Y. Chen, "Waveform contenders for 5G - suitability for short packet and low latency transmissions," in *Vehicular Technology Conference (VTC Spring), 2014 IEEE 79th*, May 2014, pp. 1–5.

[13] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 74–80, February 2014.

[14] N. DOCOMO INC, "DOCOMO 5G White paper -5G radio access: Requirements, concept and technologies," 2014.

[15] X. Jun, "Progress on 5G radio link enhancement," ZTE Technologies, Tech. Rep., June 2015.

[16] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" vol. 32, no. 6, pp. 1065–1082, June 2014.

[17] D. Kim and G. Stuber, "Residual ISI cancellation for OFDM with applications to HDTV broadcasting," vol. 16, pp. 1590–1599, 1998.

[18] C.-J. Park and G.-H. Im, "Efficient cyclic prefix reconstruction for coded OFDM systems," vol. 8, pp. 274–276, 2004.

[19] J.-H. Lee and S.-C. Kim, "Efficient ISI cancellation for STBC OFDM systems using successive interference cancellation," vol. 10, pp. 629–631, 2006.

[20] J.-B. Lim, E.-S. Kim, C.-J. Park, H.-C. Won, K.-H. Kim, and G.-H. Im, "Bandwidth-efficient OFDM transmission with iterative cyclic prefix reconstruction," *Journal of Commun. and Networks*, vol. 10, pp. 239–252, 2008.

[21] S. Ariyavisitakul and L. Greenstein, "Reduced-complexity equalization techniques for broadband wireless channels," vol. 15, pp. 5–15, 1997.

[22] S. Ariyavisitakul, N. Sollenberger, and L. Greenstein, "Tap-selectable decision feedback equalization," in *IEEE International Conf. on Commun. ICC '97 Montreal, Towards the Knowledge Millennium.*, vol. 3, Jun 1997, pp. 1521–1526 vol.3.

[23] N. Al-Dhahir and J. Cioffi, "Optimum finite-length equalization for multicarrier transceivers," *Communications, IEEE Transactions on*, vol. 44, no. 1, pp. 56–64, Jan 1996.

[24] G. Arslan, B. Evans, and S. Kiaei, "Optimum channel shortening for discrete multitone transceivers," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference on*, vol. 5, 2000.

[25] M. Nisar, W. Utschick, H. Nottensteiner, and T. Hindelang, "On channel estimation and equalization of OFDM systems with insufficient cyclic prefix," in *IEEE 65th Vehicular Techn. Conf. VTC '07-Spring.*, April 2007, pp. 1445–1449.

[26] E. Panayirci, H. Senol, and H. Poor, "Joint channel estimation, equalization, and data detection for OFDM systems in the presence of very high mobility," *Signal Processing, IEEE Transactions on*, vol. 58, no. 8, pp. 4225–4238, Aug 2010.

[27] M. Nassar, P. Schniter, and B. Evans, "A factor-graph approach to joint OFDM channel estimation and decoding in impulsive noise channels," in *Signals, Systems and Computers, 2013 Asilomar Conference on*, Nov 2013, pp. 1929–1933.

[28] M. E. Tipping and A. Smola, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, 2001.

[29] N. Pedersen, C. Navarro Manchón, D. Shutin, and B. Fleury, "Application of Bayesian hierarchical prior modeling to sparse channel estimation," in *IEEE International Conf. on Commun. ICC '12.*, June 2012, pp. 3487–3492.

[30] R. Prasad, C. Murthy, and B. Rao, "Joint Approximately Sparse Channel Estimation and Data Detection in OFDM Systems Using Sparse Bayesian Learning," vol. 62, no. 14, pp. 3591–3603, July 2014.

[31] M. Tuchler and A. Singer, "Turbo Equalization: An Overview," vol. 57, pp. 920–952, 2011.

[32] Q. Guo, L. Ping, and D. Huang, "A low-complexity iterative channel estimation and detection technique for doubly selective channels," vol. 8, pp. 4340–4349, 2009.

[33] Y. Zhu, D. Guo, and M. Honig, "A message-passing approach for joint channel estimation, interference mitigation, and decoding," vol. 8, no. 12, pp. 6008–6018, December 2009.

[34] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE Journal of Selected Topics in Signal Processing,*, vol. 5, no. 8, pp. 1462–1474, Dec 2011.

[35] C. Navarro Manchón, G. E. Kirkelund, E. Riegler, L. P. B. Christensen, and B. H. Fleury, "Receiver Architectures for MIMO-OFDM Based on a Combined VMP-SP Algorithm," *ArXiv 1111.5848*, 2011.

[36] M.-A. Badiu, C. Navarro Manchón, and B. Fleury, "Message-Passing Receiver Architecture with Reduced-Complexity Channel Estimation," vol. 17, pp. 1404–1407, 2013.

[37] E. Riegler, G. Kirkelund, C. Navarro, M. Badiu, and B. Fleury, "Merging Belief Propagation and the Mean Field Approximation: A Free Energy Approach," vol. 59, pp. 588–602, 2013.

[38] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Inc., 1988.

[39] E. P. Xing, M. I. Jordan, and S. J. Russell, "A Generalized Mean Field Algorithm for Variational Inference in Exponential Families," *ArXiv, 1212.2512*, 2012.

[40] D. Wipf and B. Rao, "Sparse Bayesian learning for basis selection," vol. 52, no. 8, pp. 2153–2164, Aug 2004.

[41] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," vol. 47, pp. 498–519, 2001.

[42] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1295–1322, June 2007.

[43] R. McEliece, D. J. C. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of Pearl's "Belief Propagation" algorithm," vol. 16, pp. 140–152, Feb 1998.

[44] T. Richardson and R. Urbanke, *Modern Coding Theory*. New York: Cambridge University Press, 2009.

[45] B. Jorgensen, *Statistical Properties of the Generalized Inverse Gaussian Distribution (Lecture Notes in Statistics 9)*. Springer-Verlang New York Inc., 1982.

[46] D. Shutin, T. Buchgraber, S. Kulkarni, and H. Poor, "Fast variational sparse Bayesian learning with automatic relevance determination for superimposed signals," vol. 59, no. 12, pp. 6257–6261, Dec 2011.

[47] N. L. Pedersen, C. N. Manchon, and B. H. Fleury, "Low complexity sparse Bayesian learning for channel estimation using generalized mean field," in *Proc of 20th European Wireless Conf. 2014;*, 2014, pp. 1–6.

[48] N. L. Pedersen, C. N. Manchón, M.-A. Badiu, D. Shutin, and B. H. Fleury, "Sparse estimation using bayesian hierarchical prior modeling for real and complex linear models," *Signal Processing*, vol. 115, pp. 94 – 109, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168415001140

[49] M. Marden, *Geometry of Polynomials*. Providence: American Mathematical Society, 1966.

# Paper B

## OFDM Receiver for Fast Time-varying Channels Using Block-sparse Bayesian Learning

Oana-Elena Barbu, Carles Navarro Manchón, Christian Rom,
Tommaso Balercia, Bernard H. Fleury

The paper has been accepted for publication in
*IEEE Transactions on Vehicular Technology*
February, 2016

# Abstract

*We propose an iterative algorithm for OFDM receivers operating over fast time-varying channels. The design relies on the assumptions that the channel response can be characterized by a few non-negligible separable multipath components, and the temporal variation of each component gain can be well characterized with a basis expansion model using a small number of terms. As a result, the channel estimation problem is posed as that of estimating a vector of complex coefficients that exhibits a block-sparse structure, which we solve with tools from block-sparse Bayesian learning. Using variational Bayesian inference, we embed the channel estimator in a receiver structure that performs iterative channel and noise precision estimation, intercarrier interference cancellation, detection and decoding. Simulation results illustrate the superior performance of the proposed receiver over state-of-art receivers.*


# 1 Introduction

OFDM systems operating over fast time-varying channels experience orthogonality loss between all system subcarrier frequencies. As a result, the observed signal is affected by intercarrier interference (ICI), which can severely degrade the receiver performance. A wide range of algorithms have been proposed in the literature to suppress or mitigate ICI [1]. They typically iterate between channel estimation, ICI cancellation and data detection [2, 3]. Among these we mention receivers implementing decision-feedback equalization (DFE), such as that proposed in [4] which performs channel estimation and ICI cancellation in the frequency domain, followed by linear equalization. Another popular category of receiver algorithms approximate the frequency-domain channel matrix to a banded version to equalize OFDM signals received over time-varying channels (see [5], [6] and the references therein). When designing channel estimators for receivers operating in such conditions, it is widespread to model the selective behavior of the channel (in time and/or frequency) by means of either: (a) an autoregressive (AR) process [2], or (b) a basis expansion [7, 8]. Estimation of the channel reduces then to estimating the model parameters, either pilot-based (by neglecting ICI) [9] or data-aided [10]. As an example of approach (a), the authors of [11] model the channel response as an AR process of low order to reflect the channel's low degrees of freedom. They apply the expectation-maximization (EM) algorithm to design a receiver algorithm that neglects ICI and iterates between hard symbol detection and channel estimation. Similar modeling is employed in [12] to design a channel estimator using approximate message passing techniques. The authors of [10] undertake the modeling approach (b) and employ space-alternating generalized EM to implement an iterative receiver algorithm which accounts for ICI. At each iteration

the algorithm computes the hard estimate of the data symbol modulating a single subcarrier, after having canceled the ICI estimated at the previous iteration. Using hard symbol estimates, and hence not accounting for uncertainties in the symbols decisions, is detrimental to the performance of the receiver when it operates over very fast time-varying channels. This is because the hard symbol decisions are highly inaccurate at the initial iteration and impair the ability of the channel estimator to refine its output at the subsequent iterations.

To overcome the above shortcomings, we propose an algorithm that iterates between estimation of the channel time-varying weights and noise precision, ICI cancellation, detection and decoding of the signals in one transmission frame. Our algorithm is developed using two main tools: block-sparse Bayesian learning (BSBL) which is a Bayesian formulation of compressed sensing [13, 14], and the mean-field belief-propagation (MF-BP) framework [15] appertaining to variational Bayesian inference. The BSBL methodology was recently applied to other communication problems such as estimating MIMO channel responses [16] or channel responses which exhibit delay clustering [17], while MF-BP was previously used for designing iterative receiver algorithms [15, Section IV], [18]. We formulate the channel estimation as a sparse signal recovery problem: the actual time-varying channel response (TV-CR) is approximated as the sum of many multipath components with delays placed on a grid of sufficiently fine resolution. The time-variant weights of these components are represented using a basis expansion model (BEM). Since typically the actual TV-CR contains only a few dominant multipath components, many components in the approximate model are zero, so that the vector collecting the BEM coefficients of all these components has many zero entries occuring in blocks, i.e. is block-sparse. We apply recovery tools from BSBL to estimate this vector, and thereby obtain an estimate of the approximate TV-CR. The channel estimator is integrated in the iterative algorithm using the MF-BP framework. In this framework, the use of soft symbol decisions results naturally from the computation of approximate posterior distributions, called beliefs. In this way, a unified iterative structure of all receiver tasks is obtained. We compare the proposed algorithm against other iterative algorithms. The first one, inspired by [11], is designed under the prevalent assumptions of negligible ICI and sparse TV-CR modeled using approach (a). The second algorithm is that proposed in [10] and thus based on approach (b). Simulation results show that the receiver implementing our algorithm outperforms selected reference algorithms and achieves BER performance close to that of a genie-aided receiver.

*Notation:* By $[P]$ we denote the set $\{p \in \mathbb{N} | 0 \leq p \leq P\}$. The matrix $\mathbf{A}$ has the $(i,j)$ entry $A[i,j]$. Its $i$th line and $j$th column are $A[i,\cdot]$ and $A[\cdot,j]$ respectively. The operator $\otimes$ designates the Kronecker product. The expected value of $f(x)$ w.r.t. the probability density function (pdf) $q(x)$ is $\langle f(x) \rangle_{q(x)}$, while the expected value of a function $f(x_1, ..., x_N)$ w.r.t. to the pfds of all its variables but one, e.g. $x_i$, is $\langle f(x_1, ..., x_N) \rangle_{q(\sim x_i)}$. We use $\mathrm{Ga}(\cdot; a, b)$ for the

gamma pdf with shape and rate parameters $a$ and $b$ respectively and $\mathrm{CN}(\cdot|\mathbf{a},\mathbf{B})$ for the multivariate complex Gaussian pdf with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{B}$.

# 2 System model

We assume an OFDM transmission of $B$ symbols. During the $i$th transmission interval, $i \in [B-1]$, a vector $\mathbf{u}_i \in \{0,1\}^K$ of information bits is encoded with a code rate $R$ and interleaved into the vector $\mathbf{c}_i = [(\mathbf{c}_i^{(0)})^{\mathrm{T}}, ..., (\mathbf{c}_i^{(N_D-1)})^{\mathrm{T}}]^{\mathrm{T}}$ with entries $\mathbf{c}_i^{(k)} \in \{0,1\}^Q, k \in [N_D-1], RN_DQ = K$. The code vector $\mathbf{c}_i$ is modulated onto a vector of $N_D$ complex symbols that are interleaved with $N_P$ pilot symbols producing the symbol vector $\mathbf{x}_i \in \mathbb{C}^N, N = N_P + N_D$. The $m$-th entry $x_i[m]$ of $\mathbf{x}_i$ is a pilot symbol if $m \in \mathcal{P}$ and a data symbol if $m \in \mathcal{D}$ [1]. The vector $\mathbf{x}_i$ is passed through an inverse DFT block to yield the vector $\mathbf{s}_i$ to which a $\mu$-sample long cyclic prefix (CP) is prepended. A frame of $B$ OFDM symbols is sent over a time-varying channel with response composed of $\tilde{L}$ multipath components: $\tilde{g}(t,\tau) = \sum_{l=0}^{\tilde{L}-1} \tilde{h}_l(t)\delta(\tau - \tilde{\tau}_l)$, where $\tilde{h}_l(t)$ and $\tilde{\tau}_l$ model the time-varying gain and delay of the $l$th multipath component[2]. The receiver observes a signal which is the convolution of the transmitted signal and the TV-CR $\tilde{g}(t,\tau)$ corrupted by additive white Gaussian noise. This signal is lowpass filtered, sampled and the CP samples are discarded. The remaining samples are collected in $B$ vectors that are passed through a DFT block, yielding

$$\mathbf{y}_i = \tilde{\mathbf{H}}_i\mathbf{x}_i + \mathbf{w}_i = \mathrm{diag}[\tilde{\mathbf{H}}_i]\mathbf{x}_i + \tilde{\mathbf{z}}_i + \mathbf{w}_i, \tag{B.1}$$

$i \in [B-1]$ and the matrix $\tilde{\mathbf{H}}_i \in \mathbb{C}^{N\times N}$ has entries $\tilde{H}_i[m,p] = \frac{1}{N}\sum_{l=0}^{\tilde{L}-1}\sum_{k=0}^{N-1} \tilde{h}_l[iN+k]\exp\left(\frac{j2\pi k(p-m)}{N} - \frac{j2\pi p\tilde{\tau}_l}{NT_s}\right)$, with $\tilde{h}_l[iN+k] = \tilde{h}_l(kT_s + i(\mu+N)T_s)$. The noise vector $\mathbf{w}_i$ has the pdf $p(\mathbf{w}_i) = \mathrm{CN}(\mathbf{w}_i; \mathbf{0}, \lambda^{-1}\mathbf{I}_N)$. The diagonal matrix $\mathrm{diag}[\tilde{\mathbf{H}}_i]$ retains the diagonal entries of $\tilde{\mathbf{H}}_i$. The vector $\tilde{\mathbf{z}}_i \in \mathbb{C}^N$ collects the ICI at all subcarriers: $\tilde{z}_i[m] = \sum_{p=0,p\neq m}^{N-1} \tilde{H}_i[m,p]x_i[p]$.

# 3 Proposed iterative receiver

## 3.1 Approximate observation model

To decode $\mathbf{u}_i$, we need first to estimate $\tilde{\mathbf{H}}_i$. Since estimating all the unknown variables involved in $\tilde{\mathbf{H}}_i$ (i.e. all $\tilde{\tau}_l$, $\tilde{h}_l[iN+k]$, $\tilde{L}$) is computationally in-

[1] Here, $\mathcal{P} = \{p_0, ..., p_{N_P-1}\}$ and $\mathcal{D} = \{d_0, ..., d_{N_D-1}\}$, $\mathcal{P}\cup\mathcal{D} = [N-1]$, $\mathcal{P}\cap\mathcal{D} = \emptyset$, are the subsets of pilot and data indices respectively, and $x_i[d_k] = \mathcal{M}(\mathbf{c}_i^{(k)})$, where $\mathcal{M}: \{1,0\}^Q \to \mathcal{S}_d$ denotes the complex modulation mapping, and $\mathcal{S}_d$ is the complex symbol constellation with cardinality $2^Q$.

[2] We assume that the maximum excess delay does not exceed the CP duration.

tractable, we seek an approximate model of this matrix whose parameters can be estimated with tractable complexity. First, we assume a grid of $L$ uniformly spaced delays [19] with selected resolution $\Delta\tau$. Then, we select the approximate model for the TV-CR: $g(t,\tau) = \sum_{l=0}^{L-1} h_l(t)\delta(\tau - l\Delta\tau)$, where $h_l(t)$ represents the postulated time-varying complex weight of the $l$th component with delay $l\Delta\tau$. Since the TV-CR of the typical wireless channel exhibits a few dominant components, we expect that, by choosing a sufficiently fine delay resolution $\Delta\tau$ and $L \gg \tilde{L}$, $g(t,\tau)$ can be used to obtain a good approximation of the matrices $\tilde{\mathbf{H}}_i$, with only a few non-zero weights $h_l(t)$ [20]. Second, we write each postulated complex gain $h_l(t)$ in $g(t,\tau)$ as a linear combination of $D$ orthonormal basis functions $\psi^{(0)}(t), ..., \psi^{(D-1)}(t)$: $h_l(t) = \sum_{d=0}^{D-1} \alpha_l[d]\psi^{(d)}(t)$ [3]. Several bases have been proposed in the literature [7, 10, 21] to model the channel time-varying behavior. Most of them are based on the Karhunen-Loeve (KL) expansion. Since the true Doppler spectrum of the channel is typically unknown at the receiver, the basis is often selected as the KL basis for an assumed Doppler spectrum, e.g. a flat or Jakes' Doppler spectrum. For each $l \in [L-1]$, we sample $h_l(t)$ at $t = (k + i(\mu+N))T_s, k \in [N-1], i \in [B-1]$ and collect the samples $h_l[iN+k] = h_l(kT_s + i(\mu+N)T_s)$ in the vector $\mathbf{h}_l \in \mathbb{C}^{NB}$. Then, $\mathbf{h}_l = \mathbf{\Psi}\boldsymbol{\alpha}_l$, where $\mathbf{\Psi}[iN+k,d] = \psi^{(d)}(kT_s + i(\mu+N)T_s)$ and $\boldsymbol{\alpha}_l = [\alpha_l[0], ..., \alpha_l[D-1]]^{\mathrm{T}}$. This formulation enables us to approximate (B.1) by

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{x}_i + \mathbf{w}_i = \mathrm{diag}[\mathbf{H}_i]\mathbf{x}_i + \mathbf{z}_i + \mathbf{w}_i, \tag{B.2}$$

where the approximate ICI vector $\mathbf{z}_i$ has entries $z_i[m] = \sum_{p=0, p\neq m}^{N-1} H_i[m,p]x_i[p]$, $m \in [N-1]$, and the approximate channel matrix $\mathbf{H}_i \in \mathbb{C}^{N\times N}$ has entries $H_i[m,p] = \frac{1}{N}\sum_{l=0}^{L-1}\sum_{k=0}^{N-1}\mathbf{\Psi}[iN+k,\cdot]\boldsymbol{\alpha}_l \exp\left(\frac{j2\pi k(p-m)}{N} - \frac{j2\pi pl\Delta\tau}{NT_s}\right)$. We collect all the vectors $\boldsymbol{\alpha}_l, l \in [L-1]$ in $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_0^{\mathrm{T}}, ..., \boldsymbol{\alpha}_{L-1}^{\mathrm{T}}]^{\mathrm{T}}$.

Note that in the model of the approximate channel matrix $\mathbf{H}_i$, for a given selection of the basis functions $\mathbf{\Psi}$, the only unknown variables are the vectors $\boldsymbol{\alpha}_l$ used to model $h_l(t)$. With this approximate model, we circumvent the explicit estimation of the multipath delays $\tilde{\tau}_l, l \in [\tilde{L}-1]$ and the number of multipath components $\tilde{L}$ in $\tilde{g}(t,\tau)$. Instead, the $DL$ entries of $\boldsymbol{\alpha}$ need to be estimated. Since only a few $h_l(t), l \in [L-1]$ are expected to be non-zero, we postulate that $\mathbf{H}_i$ can approximate $\tilde{\mathbf{H}}_i$ well with only a few non-zero vectors $\boldsymbol{\alpha}_l$. This implies that the vector $\boldsymbol{\alpha}$ will have few non-zero entries occurring in blocks of length $D$, i.e. the vector $\boldsymbol{\alpha}$ is block-sparse. Assuming a block-sparse $\boldsymbol{\alpha}$ enables the use of compressed sensing tools to retrieve its entries. Thus, we recast (B.2) as

$$\mathbf{y}_i = \mathbf{A}_i\boldsymbol{\alpha} + \mathbf{w}_i, \tag{B.3}$$

$A_i[m,\cdot] = \sum_{l=0}^{L-1}\sum_{k=0}^{N-1}\mathbf{\Omega}_i[lN+k,\cdot]\sum_{p=0}^{N-1}\frac{x_i[p]}{N}\cdot\exp\left(-2\pi j\frac{pl\Delta\tau/T_s + k(m-p)}{N}\right)$,

[3] $D \geq \lceil 2f_D NT_s B\rceil + 1$ and $D \ll NB$. Due to the page limitation we refer the reader to [10], [21] for more details on how to select $D$.

$m \in [N-1]$, $\mathbf{\Omega}_i = \mathbf{I}_L \otimes \mathbf{\Psi}_i$, and $\Psi_i[k,d] = \Psi[iN+k,d]$, $d \in [D-1]$.

## 3.2 Proposed receiver algorithm

### Probabilistic model

To enforce the block-sparse structure on $\boldsymbol{\alpha}$, we employ BSBL. BSBL is a Bayesian framework for compressed sensing which "explores and exploits the intra-block correlation" [13], i.e. makes use of the correlation between the entries of a block to retrieve block-sparse variables. This is accomplished by imposing a prior distribution for the variable of interest which encourages block-sparse estimates. We choose $p(\boldsymbol{\alpha}) = \prod_{l=0}^{L-1} p(\boldsymbol{\alpha}_l)$ with $p(\boldsymbol{\alpha}_l) = \int_0^\infty p(\boldsymbol{\alpha}_l|\gamma[l])p(\gamma[l])d\gamma[l]$ where $p(\boldsymbol{\alpha}_l|\gamma[l]) = \mathrm{CN}(\boldsymbol{\alpha}_l|\mathbf{0}, \gamma[l]^{-1}\mathbf{V})$ with $\mathbf{V}$ obtained from the Doppler power spectrum of the channel[4] [10]. For this model, $\gamma[l]$ controls the sparsity of the $l$th block, i.e. when $\gamma[l]$ takes large values, the $l$th block of $\boldsymbol{\alpha}$, namely $\boldsymbol{\alpha}_l$ is close to zero [13]. We select the sparsity inducing prior $p(\boldsymbol{\gamma}) = \prod_{l=0}^{L-1} \mathrm{Ga}(\gamma[l]; a, b)$, where the vector $\boldsymbol{\gamma}$ collects all $\gamma[l]$, $l \in [L-1]$. The selection yields $p(\boldsymbol{\alpha}|\boldsymbol{\gamma}) = \mathrm{CN}(\boldsymbol{\alpha}|\mathbf{0}, \mathbf{\Gamma}^{-1} \otimes \mathbf{V})$, where $\mathbf{\Gamma} = \mathrm{diag}(\boldsymbol{\gamma})$. The a posteriori pdf for (B.3) is $p([\mathbf{u}_b, \mathbf{c}_b, \mathbf{x}_b; b \in [B-1]], \boldsymbol{\alpha}, \boldsymbol{\gamma}, \lambda | [\mathbf{y}_b; b \in [B-1]]) \propto f_{\mathrm{n}}(\lambda) f_{\mathrm{a}}(\boldsymbol{\alpha}, \boldsymbol{\gamma}) f_{\mathrm{p}}(\boldsymbol{\gamma}) \prod_{n=0}^{B-1} f_{\mathrm{o}_n}(\mathbf{x}_n, \lambda, \boldsymbol{\alpha}) f_{\mathrm{c}_n}(\mathbf{c}_n, \mathbf{u}_n) \cdot \prod_{k=0}^{N_D-1} f_{\mathrm{m}_{n,k}}(x_n[d_k], \mathbf{c}_n^{(k)}) \prod_{v=0}^{K-1} f_{\mathrm{b}_{n,v}}(u_n[v])$, where $f_{\mathrm{n}}(\lambda) = p(\lambda) = 1/\lambda$, $f_{\mathrm{o}_n}(\mathbf{x}_n, \lambda, \boldsymbol{\alpha}) = p(\mathbf{y}_n|\mathbf{x}_n, \lambda, \boldsymbol{\alpha}) = \mathrm{CN}(\mathbf{y}_n|\mathbf{A}_n\boldsymbol{\alpha}, \lambda^{-1}\mathbf{I}_N)$, $f_{\mathrm{a}}(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = p(\boldsymbol{\alpha}|\boldsymbol{\gamma})$, $f_{\mathrm{m}}(\boldsymbol{\gamma}) = p(\boldsymbol{\gamma})$, $f_{\mathrm{b}_{n,v}}(u_n[v]) = p(u_n[v]) = 1/2$, $f_{\mathrm{m}_{n,k}}(x_n[d_k], \mathbf{c}_n^{(k)}) = \mathbb{1}_{\{\mathcal{M}(\mathbf{c}_n^{(k)})\}}(x_n[d_k])$ and $f_{\mathrm{c}_n}(\mathbf{c}_n, \mathbf{u}_n) = \mathbb{1}_{\{\mathcal{C}(\mathbf{u}_n)\}}(\mathbf{c}_n)$.[5] The two latter factors enforce respectively modulation and coding constraints.

### MF-BP

The bipartite factor graph [22] corresponding to the above probabilistic model is depicted in Fig. B.1. We define the set of variable nodes connected to the factor node $f$ as $\mathcal{N}(f)$, and similarly, the set of factor nodes connected to a variable node $\theta$ as $\mathcal{N}(\theta)$. Following the MF-BP framework we divide the set of factor nodes into two disjoint sets. The MF (BP) subgraph contains the nodes of the factor in the set $\mathcal{F}_{\mathrm{MF}} = \{f_{\mathrm{p}}, f_{\mathrm{n}}, f_{\mathrm{a}}, f_{\mathrm{o}_i}\}$ $\left(\mathcal{F}_{\mathrm{BP}} = \{f_{\mathrm{m}_{n,k}}, f_{\mathrm{c}_n}, f_{\mathrm{b}_{n,v}}\}\right)$. The messages $m^{\mathrm{BP}}_{g\to\theta}$, where $g \in \mathcal{F}_{\mathrm{BP}}$ and $\theta \in \mathcal{N}(g)$ and $m^{\mathrm{MF}}_{f\to\theta}$, where $f \in \mathcal{F}_{\mathrm{MF}}$ and $\theta \in \mathcal{N}(f)$ are updated using respectively the second and the third definitions in [15, eq. (22)]. The updates are iteratively computed until convergence or a stopping criterion is achieved. Then, the belief of any variable $\theta$ in

[4] Alternatively, $\mathbf{V}$ could be estimated [13]. Simulation results have shown, however, convergence issues; in this case, a robust selection of $\mathbf{V}$ as described in Section 4 is preferred instead.

[5] The indicator function $\mathbb{1}_A(x)$ takes value 1 if $x \in A$ and 0 otherwise.

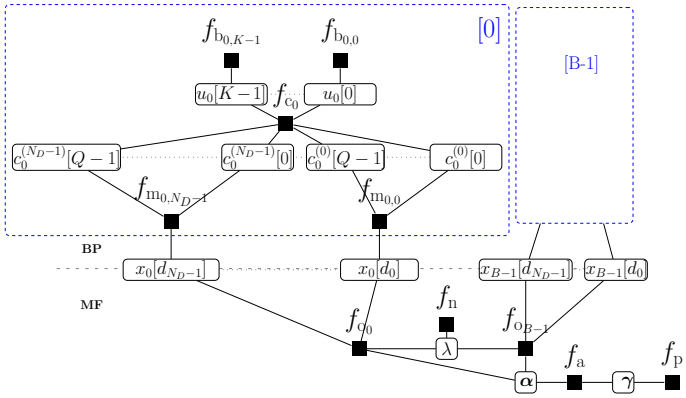


**Fig. B.1:** Factor graph of the probabilistic model.

Fig. B.1, $q(\theta) = \prod_{f \in \mathcal{N}(\theta) \cap \mathcal{F}_{\mathrm{MF}}} m^{\mathrm{MF}}_{f \to \theta}(\theta) \prod_{g \in \mathcal{N}(\theta) \cap \mathcal{F}_{\mathrm{BP}}} m^{\mathrm{BP}}_{g \to \theta}(\theta)$ approximates $p(\theta | \mathbf{y}_0, ..., \mathbf{y}_{B-1})$.

**a) Estimating the BEM coefficients:** The belief $q(\boldsymbol{\alpha}) \propto m^{\mathrm{MF}}_{f_{\mathrm{a}} \to \boldsymbol{\alpha}}(\boldsymbol{\alpha}) \prod_{n=0}^{B-1} m^{\mathrm{MF}}_{f_{\mathrm{o}n} \to \boldsymbol{\alpha}}(\boldsymbol{\alpha})$ is recast as $q(\boldsymbol{\alpha}) = \mathrm{CN}(\boldsymbol{\alpha} | \boldsymbol{\mu}_{\boldsymbol{\alpha}}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$ where

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \langle \lambda \rangle_{q(\lambda)} \, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}} \sum_{n=0}^{B-1} \langle \mathbf{A}_n \rangle^{\mathrm{H}}_{q(\mathbf{x}_n)} \, \mathbf{y}_n$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \left( \langle \lambda \rangle_{q(\lambda)} \sum_{n=0}^{B-1} \left\langle \mathbf{A}_n^{\mathrm{H}} \mathbf{A}_n \right\rangle_{q(\mathbf{x}_n)} + \langle \boldsymbol{\Gamma} \rangle_{q(\boldsymbol{\gamma})} \otimes \mathbf{V}^{-1} \right)^{-1}.$$

**b) Retrieving the block-sparse structure:** The belief $q(\boldsymbol{\gamma}) \propto m^{\mathrm{MF}}_{f_{\mathrm{a}} \to \boldsymbol{\gamma}}(\boldsymbol{\gamma}) \cdot m^{\mathrm{MF}}_{f_{\mathrm{p}} \to \boldsymbol{\gamma}}(\boldsymbol{\gamma})$ fully factors into a product of Gamma pdfs with

$$\langle \gamma[l] \rangle_{q(\gamma[l])} = (a+1) \left( b + \left\langle \boldsymbol{\alpha}_l^{\mathrm{H}} \mathbf{V}^{-1} \boldsymbol{\alpha}_l \right\rangle_{q(\boldsymbol{\alpha})} \right)^{-1}, \; l \in [L-1].$$

**c) Estimating the noise precision:** The belief $q(\lambda) \propto m^{\mathrm{MF}}_{f_{\mathrm{n}} \to \lambda}(\lambda) \prod_{n=0}^{B-1} m^{\mathrm{MF}}_{f_{\mathrm{o}n} \to \lambda}(\lambda)$ is a Gamma pdf with first moment

$$\langle \lambda \rangle_{q(\lambda)} = NB \left( \sum_{n=0}^{B-1} \left\langle \| \, \mathbf{y}_n - \mathbf{A}_n \boldsymbol{\alpha} \, \|_2^2 \right\rangle_{q(\sim \lambda)} \right)^{-1}.$$

**d) Canceling ICI and decoding:** Using (B.2) we get $q(x_i[d_k]) \propto m^{\mathrm{MF}}_{f_{\mathrm{o}_i}\to x_i[d_k]}(x_i[d_k])\; m^{\mathrm{BP}}_{f_{\mathrm{m}_{i,k}}\to x_i[d_k]}(x_i[d_k])$, where $m^{\mathrm{MF}}_{f_{\mathrm{o}_i}\to x_i[d_k]}(x_i[d_k]) = \mathrm{CN}(x_i[d_k]|\mu_{i,d_k},\rho_{i,d_k})$ with

$$\mu_{i,d_k} = \rho_{i,d_k}\,\langle\lambda\rangle_{q(\lambda)}\left(\left\langle H_i[\cdot,d_k]^{\mathrm{H}}\mathbf{y}_i\right\rangle_{q(\sim x_i[d_k])} - \sum_{j\neq k} K_i[k,j]\left\langle x_i[d_j]\right\rangle_{q(x_i[d_j])}\right)$$

$$\rho^{-1}_{i,d_k} = \langle\lambda\rangle_{q(\lambda)}\,K_i[k,k], \text{where}\quad K_i[k,j] = \left\langle H_i^{\mathrm{H}}[\cdot,d_k]H_i[\cdot,d_j]\right\rangle_{q(\sim x_i[d_k])}.$$

Note the implicit ICI cancellation: before updating $\mu_{i,d_k}$, the ICI caused by all $x_i[d_j], d_j \in \mathcal{D}, d_j \neq d_k$ is removed from the received signal $\mathbf{y}_i$, and a nearly interference-free signal (the last term in the expression of $\mu_{i,d_k}$) is employed instead. Propagating BP messages through nodes $f_{\mathrm{m}_{i,k}}$ and $f_{\mathrm{c}_i}$ corresponds to classical demapping and decoding respectively [23], hence we leave out their updates. After the beliefs $q(x_i[d_k])$ are computed, $\langle\mathbf{A}_i\rangle_{q(\mathbf{x}_i)}$ and $\left\langle\mathbf{A}_i^{\mathrm{H}}\mathbf{A}_i\right\rangle_{q(\mathbf{x}_i)}$ are updated. The proposed algorithm is set with fixed values for $\Delta\tau, L$ and $\mathbf{\Psi}$. It initializes $\langle\lambda\rangle_{q(\lambda)}$, $\langle\boldsymbol{\gamma}\rangle_{q(\boldsymbol{\gamma})}$ and $\langle\mathbf{x}_i\rangle_{q(\mathbf{x}_i)}, \left\langle|\mathbf{x}_i|^2\right\rangle_{q(\mathbf{x}_i)}$, $i \in [B-1]$, and then iterates among the updates *a) – d).*

# 4 Numerical evaluation

We generate realizations of the TV-CR $\tilde{g}(t,\tau)$ from Section 2 as follows. The number of multipath components $\tilde{L}$ is drawn from a Poisson distribution with mean $\mu_{\tilde{L}}$. Given $\tilde{L}$, to each component $l \in [\tilde{L}-1]$ a random vector $(\tilde{\tau}_l, z_l, \varphi_l, [\vartheta_{l,k}, \varsigma_{l,k}; k \in [M-1]])$ is associated: the delay $\tilde{\tau}_l$ is uniformly distributed on $[0, \mu T_s]$; given $\tilde{\tau}_l$, the gain $z_l$ is a zero-mean complex Gaussian variable with variance $v_0\exp(-m_0\tau_l)$; the mean azimuth $\varphi_l$ and the phases $\varsigma_{l,k}$ are drawn from a uniform distribution on $[0, 2\pi)$. Given $\varphi_l$, the azimuths $\vartheta_{l,k}$ are drawn from a von Mises distribution with mean $\varphi_l$ and concentration $\kappa_v$. Consequently, the time-varying gain of the $l$th multipath component reads $\tilde{h}_l(t) = z_l \sum_{n=0}^{M-1}\exp[j(2\pi f_D t\cos\vartheta_{l,n} + \varsigma_{l,n})]$. Table B.2 lists the system and channel parameter settings.

We abbreviate the receiver implementing the proposed algorithm as RxBEM-sparse. We obtain $\mathbf{V}$ by assuming Jakes' Doppler power spectrum and we choose the discrete cosine orthogonal basis functions [10]. We set $\Delta\tau = 100ns$, $L = \lceil\mu T_s/\Delta\tau\rceil$ and $a = 2, b = 0.1$ and assume known $f_D$. We initialize $\langle\lambda\rangle_{q(\lambda)} = 1$, $\langle\boldsymbol{\gamma}\rangle_{q(\boldsymbol{\gamma})} = \mathbf{1}$. To assess the benefits of jointly exploiting the channel sparsity and iteratively canceling ICI we compare RxBEM-sparse against the benchmark receivers from Table B.1. For channel estimation benchmarking we use a genie-aided channel estimator (GAE) which performs LMMSE estimation of $\boldsymbol{\alpha}$ in (B.3) with known dictionary matrix[6] and noise variance. In

[6] The dictionary matrix $\mathbf{A}_i$ is constructed under the assumption that the modulated

**Table B.1:** Benchmark receivers

| Benchmarker | Implementation |
|---|---|
| RxAR (based on [11])<br>- neglects ICI | - Algorithm in Appendix with a sparse channel estimator. |
| RxRef-ss<br>- cancels ICI iteratively<br>- estimates hard symbols | - Alg. from [10] enhanced with iterative decoding.<br>- knows the noise variance.<br>- assumes $T_s$-spaced delays. |
| RxRef (see RxRef-ss) | - assumes $\Delta\tau$-spaced delays. |
| GAR (genie)<br>- cancels ICI perfectly | - perfect channel information.<br>- knows the noise variance. |

**Table B.2:** Setting of the parameters for the simulations.

| System | TV-CR |
|---|---|
| $N = 100, \mu = 7, N_P = 17$ | $1/m_0 = 3\mu s$ |
| $B = 10,\ f_c = 2.6$ GHz | $M = 3, \kappa_v = 10$ |
| $f_s = 15$ kHz, $T_s = (Nf_s)^{-1}$, | $\mu_{\bar{L}} = 10$ |
| QPSK , 1/3 code rate | $f_D = 1.2$ kHz |

Fig. B.2 we observe that the average MSE for all receivers exhibits a saturation when they operate in the high SNR regime. This behavior is due to the estimation model mismatch stemming from the choice of basis, the fixed delay grid and the errors in the estimates of the data symbols. In particular, RxRef–ss exhibits the highest sensitivity to these mismatches as it employs a delay vector with $T_s$-spaced entries. Even though there is still a notable gap between the average MSEs of RxBEM-sparse and GAE, RxBEM-sparse performs very closely to GAR in terms of BER outperforming all benchmark receivers. This shows that efficient receivers can accommodate some errors in the estimation of the channel and still operate closely to the optimal performance. Canceling ICI using hard symbol estimates proves detrimental to the BER performance of both RxRef and RxRef–ss, particularly in the high SNR regime. In this case, an algorithm which neglects ICI, such as RxAR is preferred.

# 5 Conclusion

To develop a tractable algorithm for OFDM receivers operating over fast time-varying channels, we employ the BEM method to model the TV-CR weights. Assuming a finely sampled delay support enables a design which harnesses the block-sparse structure of the resulting BEM vector and thus overcomes the task of explicitly estimating the delays. In the studied conditions, a receiver

symbols, the concentration parameter, the mean azimuth and the delay of each multipath component are known.

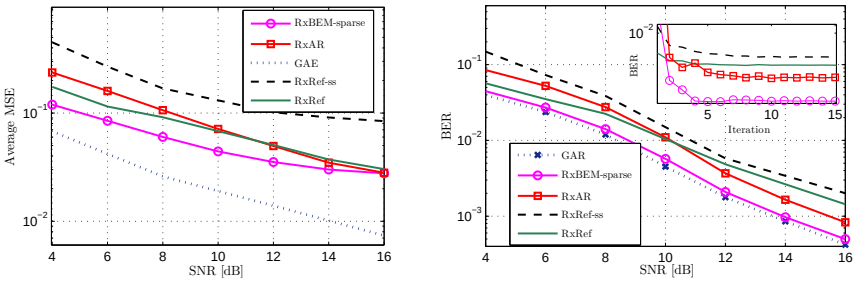


**Fig. B.2:** Average MSE of the channel frequency response versus SNR (left), BER versus SNR (right) and BER versus iteration index at 12 dB SNR (right insert).

implementing our algorithm successfully cancels ICI and performs closely to a genie-aided receiver.

# A Algorithm derived using [11]

We assume block fading $\beta_i[l] = h_l\left(i(\mu+N)T_s\right)$ and use [11]: $\boldsymbol{\beta}_i = \rho\boldsymbol{\beta}_{i-1} + \mathbf{v}_i$, $p(v_i[l]|\theta[l]) = \mathrm{CN}(v_i[l]|0,(1-\rho^2)\theta[l])$ , $p(\theta[l]) = \mathrm{Ga}(\theta[l];a,b)$.

**Algorithm** Algorithm derived using [11] and MF-BP.

1: set $L, \Delta\tau$ and init. $\langle\lambda\rangle_{q(\lambda)}$, $\langle\boldsymbol{\theta}\rangle_{q(\boldsymbol{\theta})}$, $\langle\mathbf{x}_i\rangle_{q(\mathbf{x}_i)}$, $\left\langle|\mathbf{x}_i|^2\right\rangle_{q(\mathbf{x}_i)}$
2: for $i_{out} = [1:O_M]$
3: for $i_{in} = [1:I_M]$ and $i \in [B-1]$
4: $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_i}^{-1} = \left\langle \lambda \parallel \mathbf{E}_i \parallel_2^2 + \frac{(\mathrm{diag}(\boldsymbol{\theta}))^{-1}}{(1-\rho^2)^{\mathrm{sgn}(i)}} \right\rangle_{q(\sim\boldsymbol{\beta}_i)}$,
5: $\boldsymbol{\mu}_{\boldsymbol{\beta}_i} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i}\left\langle \lambda\mathbf{E}_i^{\mathrm{H}}\mathbf{y}_i + \frac{\rho^{\mathrm{sgn}(i)}}{(1-\rho^2)^{\mathrm{sgn}(i)}}(\mathrm{diag}(\boldsymbol{\theta}))^{-1}\ \boldsymbol{\mu}_{\boldsymbol{\beta}_{i-1}} \right\rangle_{q(\sim\boldsymbol{\beta}_i)}$
6: $\langle\lambda\rangle_{q(\lambda)} = NB\left(\sum_{n=0}^{B-1}\left\langle \parallel \mathbf{y}_n - \mathbf{E}_n\boldsymbol{\beta}_n \parallel_2^2 \right\rangle_{q(\sim\lambda)}\right)^{-1}$
7: $\left\langle\theta[l]^{-1}\right\rangle_{q(\theta[l])} = \left(\frac{b}{S_B[l]}\right)^{\frac{1}{2}} \frac{K_{a-2}(2\sqrt{bS_B[l]})}{K_{a-1}(2\sqrt{bS_B[l]})}$
8: end for
9: $\langle x_i[d_k]\rangle_{q(x_i[d_k])}$, $\left\langle|x_i[d_k]|^2\right\rangle_{q(x_i[d_k])} \leftarrow d)$, $d_k \in \mathcal{D}$
10: end for

$\rho = J_0(2\pi f_D(\mu+N)T_s)$, $E_i[p,l] = x_i[p]\exp(-j2\pi pl\Delta\tau f_s)$

$S_B[l] = \dfrac{\sum_{n=0}^{B-1}\left(\left\langle|\beta_n[l]|^2\right\rangle_{q(\boldsymbol{\beta}_n)} + \left\langle|\beta_{n-1}[l]|^2\right\rangle_{q(\boldsymbol{\beta}_{n-1})} - 2\Re\{\rho\}\right)}{(1-\rho^2)^{\mathrm{sgn}(n)}}$

$J_0(x)$ is the Bessel function of the first kind and order 0. $K_p(x)$ is the modified Bessel functions of the second kind, and order $p$.

# References

[1] A. Molisch, M. Toeltsch, and S. Vermani, "Iterative methods for cancellation of intercarrier interference in OFDM systems," vol. 56, no. 4, 2007.

[2] H. Hijazi and L. Ros, "Joint data QR-detection and Kalman estimation for OFDM time-varying Rayleigh channel complex gains," vol. 58, no. 1, 2010.

[3] S. Tomasin, A. Gorokhov, H. Yang, and J.-P. Linnartz, "Iterative interference cancellation and channel estimation for mobile OFDM," vol. 4, no. 1, 2005.

[4] A. Gorokhov and J.-P. Linnartz, "Robust OFDM receivers for dispersive time-varying channels: equalization and channel acquisition," vol. 52, no. 4, 2004.

[5] W. G. Jeon, K. H. Chang, and Y. S. Cho, "An equalization technique for orthogonal frequency-division multiplexing systems in time-variant multipath channels," vol. 47, no. 1, Jan 1999.

[6] P. Schniter, "Low-complexity equalization of OFDM in doubly selective channels," vol. 52, 2004.

[7] P. Salvo Rossi and R. Muller, "Joint iterative time-variant channel estimation and multi-user detection for MIMO-OFDM systems," in *Proc. IEEE GLOBECOM*, 2007.

[8] P. Cheng, Z. Chen, Y. Rui, Y. Guo, L. Gui, M. Tao, and Q. Zhang, "Channel estimation for OFDM systems over doubly selective channels: A distributed compressive sensing based approach," vol. 61, no. 10, 2013.

[9] C. Chen and M. Zoltowski, "Bayesian sparse channel estimation and tracking," in *Proc. IEEE SSP*, 2012.

[10] E. Panayirci, H. Senol, and H. Poor, "Joint channel estimation, equalization, and data detection for OFDM systems in the presence of very high mobility," 2010.

[11] R. Prasad, C. Murthy, and B. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse Bayesian learning," 2014.

[12] J. Ziniel, L. Potter, and P. Schniter, "Tracking and smoothing of time-varying sparse signals via approximate belief propagation," in *Proc. Asilomar conf.*, Nov 2010.

[13] Z. Zhang and B. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE Journal of Sel. Topics in Sig. Processing*, vol. 5, no. 5, Sept 2011.

[14] D. Wipf and B. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," vol. 55, no. 7, July 2007.

[15] E. Riegler, G. Kirkelund, C. Navarro, M. Badiu, and B. Fleury, "Merging belief propagation and the mean field approximation: a free energy approach," vol. 59, 2013.

[16] R. Prasad, C. Murthy, and B. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," vol. 63, 2015.

[17] G. Gui, L. Xu, and L. Shan, "Block Bayesian sparse learning algorithms with application to estimating channels in OFDM systems," in *Proc. WPMC*, 2014.

[18] M.-A. Badiu, C. Navarro Manchón, and B. Fleury, "Message-Passing Receiver Architecture with Reduced-Complexity Channel Estimation," vol. 17, 2013.

[19] N. Pedersen, C. Manchon, D. Shutin, and B. Fleury, "Application of Bayesian hierarchical prior modeling to sparse channel estimation," in *IEEE ICC*, June 2012, pp. 3487–3492.

[20] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, June 2010.

[21] T. Zemen and C. Mecklenbräuker, "Time-variant channel estimation using discrete prolate spheroidal sequences," vol. 53, no. 9, 2005.

[22] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," 2001.

[23] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, no. 6, 2007.

# Paper C

## Interference-aware OFDM Receiver for Channels with Sparse Common Supports

Oana-Elena Barbu, Carles Navarro Manchón, Mihai-Alin Badiu, Christian Rom, Tommaso Balercia, Bernard H. Fleury

The paper has been submitted to
*IEEE Global Communications Conference, GLOBECOM*
March, 2016

# Abstract

*We design an algorithm for MIMO OFDM receivers operating in co-channel interference conditions using Bayesian inference. The channel estimation problem is formulated as one of sparse signal reconstruction using multiple measurement vectors. What sets our work apart is that the proposed design harnesses the sparse common support of, respectively, the desired and interfering MIMO sub-channels by adopting a Bernoulli-Gaussian prior for the weights of the impulse responses of these sub-channels. Then, applying a variational Bayesian inference method we derive an algorithm that performs joint channel estimation, interference cancellation and decoding. Simulation results show how the performance of the proposed receiver depends on its knowledge of the interfering signals' modulation and code. When these are known, our receiver approaches the performance of a genie-aided receiver with perfect interference cancellation. Even with mismatched assumptions on the modulation and code, our proposed implementation still outperforms receivers which neglect the interference.*


# 1 Introduction

Spectrum sharing and network densification, two key features of 5G networks, make the co-channel interference (CI) problem increasingly stringent [1] in the downlink. Favored by a device-centric 5G network architecture [2], algorithms for intelligent MIMO receivers (i.e. receivers with increased computational power capable of performing interference rejection/cancellation) became increasingly popular in recent years.

We can group the interference-aware receivers into two categories. The first category includes receivers implementing algorithms that suppress CI by performing interference rejection combing (IRC) [3], [4] under the assumption that accurate channel estimates are available. The second category contains more advanced receivers that implement various interference cancellation techniques to estimate and remove CI. Several algorithms in this category exploit the available information about the interferers and perform desired and interfering symbol detection and channel estimation. Prior to detecting the symbols in an interfering signal, however, the receiver needs to obtain its modulation alphabet, either by getting access to the control channels carrying this information (not possible with the current system design), by estimating it, or by assuming a fixed alphabet, at the risk of this assumption being mismatched [5]. Receivers from both categories face a common challenge: the channel state information (CSI) for all – desired and interfering – MIMO sub-channels has to be acquired. This is typically accomplished by estimating the sub-channel responses using the reference signals, the so-called pilots, of each transmitter. When the pilots are interference-free (a condition however difficult to ensure in

practice) accurate CSI can be computed [1, 3]. However, without cooperation, data symbols from transmitters simultaneously utilizing the same spectrum will interfere with each others' pilots. A pilot-based channel estimator therefore yields poor-quality CSI that further impairs the equalization and causes an overall degradation of the receiver's performance. Hence, there is a need for algorithms which account for the interference both when acquiring CSI and when equalizing the desired signal. Such alternative is proposed in [6]. The authors present iterative schemes for soft symbol detection and estimation of the frequency responses of the sub-channels under the assumption that their prior covariance matrix is known. However, as the channel covariance matrix is typically unknown, some degradation of the overall performance is expected.

Contrary to the above approaches, our design embeds an estimator of the impulse responses of the MIMO sub-channels that exploits two assumed properties of these responses, namely that they are sparse [7] and those associated with the same transmit antenna have a common support [8]. The first assumption is motivated by the fact that, in many propagation environments, the impulse response of the channel (including a transmit antenna and a receive antenna) consists of only a few significant multipath components. The second assumption is motivated by the fact that the receive antennas are co-located. As a result, the sets of delays of the significant components of the impulse responses associated with the same transmit antenna are identical. Their amplitude may, however, differ. We say therefore that these responses exhibit a sparse common support (SCS) [9]. Under these two assumptions we pose the channel estimation problem as one of reconstructing sparse signals from multiple measurement vectors (MMV) that share the same unknown sparse support [10] and employ a Bernoulli-Gaussian prior model for the weights of the MIMO sub-channels. To the authors' knowledge, such model has not been applied before to design receiver algorithms tailored for operating in co-channel interference conditions. Note that our receiver requires only prior knowledge of the maximum excess delay of the channels, in contrast to [6] which assumes knowledge of the channel's frequency autocorrelation function. To embed the channel estimator in a joint receiver design, we apply the variational Bayesian inference method combining mean-field and belief propagation [11]. The method was used successfully in the recent years to design algorithms that perform joint channel estimation and decoding [12, 13].

We test our receiver against the receiver in [6] and a genie-aided receiver based on [4] and analyze its sensitivity to mismatched information about the interfering signals in various signal-to-interference (SIR) regimes. The numerical evaluation shows that, even though the pilots of different transmitters overlap, our receiver successfully exploits the SCS property and approaches the performance of the genie-aided receiver. Knowledge of the interferers' modulation and code yields significant performance gains, which indicates that enabling access to this information could be an advantageous feature for future 5G sys-

tems.

*Notation:* By $[N:P]$ we denote the set $\{p \in \mathbb{N} |\ N \leq p \leq P\}$ and by $[P]$ the set $[1:P]$. The matrix $\mathbf{A}$ has the $(i,j)$ entry $A[i,j]$, while its $i$th row and $j$th column are $\mathbf{A}[i,\cdot]$ and $\mathbf{A}[\cdot,j]$ respectively; $\mathrm{vec}(\mathbf{A})$ returns a column vector consisting of all columns of $\mathbf{A}$ stacked on top of one another, and $\mathbf{A}[r_1 : r_N, \cdot] = [\mathbf{A}[r_1,\cdot]^{\mathrm{T}} \dots \mathbf{A}[r_N,\cdot]^{\mathrm{T}}]^{\mathrm{T}}$. The diagonal matrix $\mathbf{A} = \mathrm{diag}(\mathbf{a})$ has the entries of the vector $\mathbf{a}$ on its diagonal. The operator $\otimes$ designates the Kronecker product. The expected value of $f(x_1, ..., x_N)$ w.r.t. to the pdf of $x_k$ is $\langle f(x_1, ..., x_N)\rangle_{x_k}$, while the expected value w.r.t to the pfds of all its variables but $x_i$, is $\langle f(x_1, ..., x_N)\rangle_{\sim x_i}$. We use $\mathrm{CN}(\cdot|\mathbf{a}, \mathbf{B})$ for the multivariate complex Gaussian probability density function (pdf) with mean vector $\mathbf{a}$ and covariance matrix $\mathbf{B}$, $\mathrm{Bern}(\cdot|p)$ for the Bernoulli distribution with parameter $p \in [0,1]$, $\mathrm{U}(\cdot|a,b)$ for the uniform pdf in the interval $[a,b]$, and $\mathrm{Poiss}(\cdot|\mu)$ for the Poisson distribution with parameter $\mu > 0$. We define the indicator function $\mathbb{1}_{\mathcal{S}}(x) = 1, \text{if } x \in \mathcal{S}$ and $\mathbb{1}_{\mathcal{S}}(x) = 0, \text{otherwise}$. We use the notation $d(\mathbf{a}, \mathbf{b})$ for the Hamming distance between the vectors $\mathbf{a}$ and $\mathbf{b}$ and $\Re\{x\}$ for the real part of $x$.

# 2 Signal model

We consider a MIMO system with $T$ antennas synchronously transmitting OFDM signals and a receiver equipped with $R$ antennas. We identify a transmit (receive) antenna with its index $t \in [T]$ ($r \in [R]$). Similarly we identify the sub-channel including transmit antenna $t$, the propagation environment, and receive antenna $r$ with the pair $(t,r)$. We assume that the antennas in $[U] \subseteq [T], 1 \leq U \leq T$, send signals intended to the receiver, while the remaining $I = T - U$ antennas $[U+1, T]$ send interfering signals. A transmitter $t \in [T]$ generates a vector $\mathbf{u}^{(t)} \in \{0,1\}^K$ of information bits which it encodes with a code rate $C$ to the vector $\mathbf{c}^{(t)} = \mathcal{C}(\mathbf{u}^{(t)}) = [(\mathbf{c}_1^{(t)})^{\mathrm{T}}, ..., (\mathbf{c}_{N_{\mathrm{d}}}^{(t)})^{\mathrm{T}}]^{\mathrm{T}} \in \{0,1\}^{K/C}$, where $\mathbf{c}_k^{(t)} \in \{0,1\}^Q, K = N_{\mathrm{d}}QC$, and $\mathcal{C}$ denotes the coding and interleaving mapping. This codeword is mapped to the vector $\mathbf{x}_{\mathrm{d}}^{(t)} \in \mathbb{C}^{N_{\mathrm{d}}}$ with entries $x_{\mathrm{d}}^{(t)}[k] = \mathcal{M}(\mathbf{c}_k^{(t)}), k \in [N_{\mathrm{d}}]$, with $\mathcal{M}$ denoting the complex modulation mapping. The symbols are interleaved with $N_{\mathrm{p}}$ pilots from the vector $\mathbf{x}_{\mathrm{p}}^{(t)}$ to yield the vector $\mathbf{x}^{(t)} \in \mathbb{C}^N, N = N_{\mathrm{d}} + N_{\mathrm{p}}$ of transmitted symbols. The $i_k \in [N]$ entry of $\mathbf{x}^{(t)}$ is $x^{(t)}[i_k] = x_{\mathrm{d}}^{(t)}[k] \cdot \mathbb{1}_{\mathcal{D}^{(t)}}(i_k) + x_{\mathrm{p}}^{(t)}[k] \cdot \mathbb{1}_{\mathcal{P}^{(t)}}(i_k)$, where $\mathcal{D}^{(t)}$ ($\mathcal{P}^{(t)}$) is the subset of $N_{\mathrm{d}}$ data ($N_{\mathrm{p}}$ pilot) indices of transmitter $t$, and $\mathcal{D}^{(t)} \cap \mathcal{P}^{(t)} = \emptyset$. The vector $\mathbf{x}^{(t)}$ is then modulated to yield an OFDM waveform, including a cyclic prefix (CP) of duration $d_{\mathrm{c}}$ which is sent across the wireless channel. We

model the impulse response of sub-channel $(t, r)$ as

$$g^{(t,r)}(\tau) = \sum_{m \in [L^{(t)}]} h^{(t,r)}[m] \delta \left( \tau - \tau^{(t)}[m] \right), \tag{C.1}$$

where $h^{(t,r)}[m]$ and $\tau^{(t)}[m]$ are the weight and delay[1] of the $m$th multipath component. These parameters are collected in the vectors $\mathbf{h}^{(t,r)} \in \mathbb{C}^{L^{(t)}}$ and $\boldsymbol{\tau}^{(t)} \in \mathbb{R}^{L^{(t)}}$ respectively. In (C.1) we assumed that the receive antennas are closely spaced so that for each $t \in [T]$, the impulse responses of the sub-channels $(r, t)$ $r \in [R]$ consist of the same number $L^{(t)}$ of multipath components with the same delay vector $\boldsymbol{\tau}^{(t)}$. The signal received at antenna $r$ is sampled and OFDM demodulated (after removing the CP). The resulting vector of samples $\mathbf{y}^{(r)} \in \mathbb{C}^N$ is given by

$$\mathbf{y}^{(r)} = \sum_{t \in [T]} \mathbf{X}^{(t)} \mathbf{D}^{(t)} \mathbf{h}^{(t,r)} + \boldsymbol{\xi}^{(r)}, \tag{C.2}$$

where $D^{(t)}[n, m] = \exp\left( -j2\pi(n-1)\Delta f \tau^{(t)}[m] \right)$, $n \in [N]$, $m \in [L^{(t)}]$, $\Delta f$ is the subcarrier spacing, $\mathbf{X}^{(t)} = \operatorname{diag}\left( \mathbf{x}^{(t)} \right)$ and the noise vector has the pdf $p(\boldsymbol{\xi}^{(r)}) = \mathrm{CN}(\boldsymbol{\xi}^{(r)} | 0, \lambda^{-1} \mathbf{I}_N)$.

# 3 Message-passing receiver design

## 3.1 Estimation model

Instead of estimating $L^{(t)}$ and $\boldsymbol{\tau}^{(t)}$ needed to reconstruct the channel frequency responses $\mathbf{D}^{(t)} \mathbf{h}^{(t,r)}$ in (C.2), we assume the delays to be restricted to a grid $\bar{\boldsymbol{\tau}} \in \mathbb{R}^{L_\Delta}$ with entries $\bar{\tau}[k] = (k-1)\Delta\tau, k \in [L_\Delta]$ where $\Delta\tau$ is the chosen delay resolution. This enables us to define a dictionary $\bar{\mathbf{D}} \in \mathbb{C}^{N \times L_\Delta}$ with entries $\bar{D}[n, k] = \exp\left( -j2\pi(n-1)\Delta f (k-1)\Delta\tau \right), n \in [N], k \in [L_\Delta]$, and a vector $\bar{\mathbf{h}}^{(t,r)} \in \mathbb{C}^{L_\Delta}$ for every $t \in [T]$ and $r \in [R]$ so that $\bar{\mathbf{D}} \bar{\mathbf{h}}^{(t,r)} \approx \mathbf{D}^{(t)} \mathbf{h}^{(t,r)}$. For this approximation to be accurate, a fine enough resolution $\Delta\tau$ and a large value $L_\Delta \gg \max(L^{(1)}, .., L^{(T)})$ need to be selected. With these choices, we expect that the channel frequency responses can be well approximated using vectors $\bar{\mathbf{h}}^{(t,r)}$ having many zero entries, i.e. being sparse [14].

Then, we approximate (C.2) by $\mathbf{y}^{(r)} = \sum_{t \in [T]} \mathbf{X}^{(t)} \bar{\mathbf{D}} \bar{\mathbf{h}}^{(t,r)} + \boldsymbol{\xi}^{(r)}$ and cast the MMV $\mathbf{Y} = \bar{\boldsymbol{\Phi}} \bar{\mathbf{H}} + \boldsymbol{\Xi}$, where we define the matrices $\mathbf{Y} = [\mathbf{y}^{(1)} \dots \mathbf{y}^{(R)}]$, $\boldsymbol{\Xi} = [\boldsymbol{\xi}^{(1)} \dots \boldsymbol{\xi}^{(R)}]$, $\bar{\boldsymbol{\Phi}} = [\mathbf{X}^{(1)} \bar{\mathbf{D}} \dots \mathbf{X}^{(T)} \bar{\mathbf{D}}] \in \mathbb{C}^{N \times \bar{L}}$, and $\bar{L} = T L_\Delta$. The matrix $\bar{\mathbf{H}} \in \mathbb{C}^{\bar{L} \times R}$ has block-entries $\bar{\mathbf{H}}[(t-1)L_\Delta + 1 : tL_\Delta, \cdot] = [\bar{\mathbf{h}}^{(t,1)} \dots \bar{\mathbf{h}}^{(t,R)}], t \in$

[1] We assume that for each $t \in [T]$, the maximum excess delay does not exceed the CP duration.

$[T]$. Given the small spatial separation of the receive antennas we assume that the vectors $\bar{\mathbf{h}}^{(t,1)}, ..., \bar{\mathbf{h}}^{(t,R)}$ have the same sparsity pattern, i.e. they have a common sparse support. This implies that $\bar{\mathbf{H}}$ has many zero rows, i.e. $\bar{\mathbf{H}}$ is row-sparse. We exploit this property to estimate $\bar{\mathbf{H}}$ as follows. We model each row of $\bar{\mathbf{H}}$ as $\bar{\mathbf{H}}[l,\cdot] = s[l]\mathbf{v}_l^{\mathrm{T}}, l \in [\bar{L}]$, where $\mathbf{v}_l \in \mathbb{C}^R$ is the complex weight vector of the $l$th component, and $s[l] \in \{0,1\}$ is the $l$th entry of the binary vector $\mathbf{s}$. The vector $\mathbf{s}$ gives the delay support of all sub-channels, i.e. $s[l] = 0$ sets $\bar{\mathbf{H}}[l,\cdot] = \mathbf{0}$. We choose $p(\mathbf{v}_l) = \mathrm{CN}(\mathbf{v}_l|\mathbf{0}, \mathbf{K})$ with $\mathbf{K} \in \mathbb{C}^{R\times R}$ known. We model $p(s[l]|\zeta) = \mathrm{Bern}(s[l]|\zeta)$, and $p(\zeta) = \mathrm{U}(\zeta|0,1)$. We re-write the MMV as

$$\mathbf{y} = \mathbf{\Omega}\mathrm{vec}(\bar{\mathbf{H}}) + \mathrm{vec}(\mathbf{\Xi}) \tag{C.3}$$

where $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$ and $\mathbf{\Omega} = \mathbf{I}_R \otimes \bar{\mathbf{\Phi}}$. The posterior pdf of all latent variables in the model, given the observation (C.3) reads

$$\begin{aligned} &p\left([\mathbf{v}_l; l \in [\bar{L}]], [\mathbf{x}_{\mathrm{d}}^{(t)}, \mathbf{c}^{(t)}, \mathbf{u}^{(t)}; t \in [T]], \mathbf{s}, \zeta, \lambda|\mathbf{y}\right) \propto \\ &p(\mathbf{y}|\mathbf{s}, \lambda, [\mathbf{v}_l; l \in [\bar{L}]], [\mathbf{x}_{\mathrm{d}}^{(t)}; t \in [T]]) p(\lambda) \cdot \\ &p(\zeta) \prod_{l=1}^{\bar{L}} p(s[l]|\zeta) p(\mathbf{v}_l) \prod_{t\in[T]} \mathbb{1}_{\{\mathcal{C}(\mathbf{u}^{(t)})\}}(\mathbf{c}^{(t)}) \cdot \\ &\prod_{k=1}^{N_D} \mathbb{1}_{\{\mathcal{M}(\mathbf{c}_k^{(t)})\}}(x_{\mathrm{d}}^{(t)}[k])) \prod_{j=1}^{K} p(u^{(t)}[j]), \end{aligned} \tag{C.4}$$

where the first factor is the likelihood $p(\mathbf{y}|\mathbf{s}, \lambda, [\mathbf{v}_l; l \in [\bar{L}]], [\mathbf{x}_{\mathrm{d}}^{(t)}; t \in [T]]) = \mathrm{CN}(\mathbf{y}|\mathbf{\Omega}\mathrm{vec}(\bar{\mathbf{H}}), \lambda^{-1}\mathbf{I}_{NR})$. The factors $\mathbb{1}_{\{\mathcal{M}(\mathbf{c}_k^{(t)})\}}(x_{\mathrm{d}}^{(t)}[k]))$ and $\mathbb{1}_{\{\mathcal{C}(\mathbf{u}^{(t)})\}}(\mathbf{c}^{(t)}), t \in [T], k \in [N_{\mathrm{d}}]$ enforce modulation and coding constraints. We assume $p(u^{(t)}[j]) = 1/2, t \in [T], j \in [K]$ and select a non-informative prior for the noise precision, i.e. $p(\lambda) \propto 1/\lambda$.

**Table C.1:** Definition of the factors in (C.4)

| |
|---|
| $f_{\mathrm{n}}(\lambda) = p(\lambda)$ |
| $f_{\mathrm{a}}(\zeta) = p(\zeta)$ |
| $f_{\mathrm{h}_l}(\mathbf{v}_l) = p(\mathbf{v}_l), l \in [\bar{L}]$ |
| $f_{\mathrm{s}}(\mathbf{s}, \zeta) = \prod_{l=1}^{\bar{L}} p(s[l]\|\zeta)$ |
| $f_{\mathrm{o}}(\mathbf{s}, \lambda, [\mathbf{v}_l; l \in [\bar{L}]], [\mathbf{x}_{\mathrm{d}}^{(t)}; t \in [T]]) =$ |
| $= p(\mathbf{y}\|\mathbf{s}, \lambda, [\mathbf{v}_l; l \in [\bar{L}]], [\mathbf{x}_{\mathrm{d}}^{(t)}; t \in [T]])$ |
| $f_{\mathrm{u}_j^{(t)}}(u^{(t)}[j]) = p(u^{(t)}[j]), j \in [K]$ |
| $f_{\mathrm{c}^{(t)}}(\mathbf{c}^{(t)}, \mathbf{u}^{(t)}) = \mathbb{1}_{\{\mathcal{C}(\mathbf{u}^{(t)})\}}(\mathbf{c}^{(t)}), t \in [T]$ |
| $f_{\mathrm{m}_k^{(t)}}(x_{\mathrm{d}}{}^{(t)}[k], \mathbf{c}_k^{(t)}) = \mathbb{1}_{\{\mathcal{M}(\mathbf{c}_k^{(t)})\}}(x_{\mathrm{d}}{}^{(t)}[k]), k \in [N_{\mathrm{d}}]$ |

We collect the functions defined in Table C.1 in the set $\mathcal{F}$ and cast the factor graph of (C.4) in Fig. C.1.

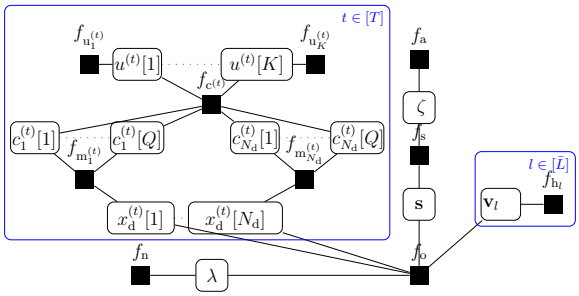


**Fig. C.1:** Factor graph representation of (C.4).

## 3.2 Algorithm description

To design an algorithm that iterates between channel, noise and interference estimation, equalization and decoding we apply the MF-BP inference framework [11], that combines belief propagation (BP) [15] and mean-field approximation (MF) [16]. The application of this method to joint channel estimation and decoding is presented in [11, Section IV]. Using MF-BP we compute an approximation (also called belief) $q(z)$ of the posterior probability density of any unobserved variable $z \in \{\mathbf{v}_l; l \in [\bar{L}]\} \cup \left\{\mathbf{x}_\mathrm{d}^{(t)}, \mathbf{c}^{(t)}, \mathbf{u}^{(t)}; t \in [T]\right\} \cup \{\mathbf{s}, \zeta, \lambda\}$ by iteratively exchanging messages on the factor graph. We denote a message from factor node $f$ to the neighboring variable node $z$ as $m_{f\to z}$ and the message in the opposite direction as $n_{z\to f}$ and we compute their updates using [11, eq. (22)]. The belief $q(z)$ is computed using [11, eq. (21)] as the product of all messages sent to the variable node $z$. Following [11], we define a MF and a BP subgraph of the factor graph in Fig. C.1. To that end, we split the set $\mathcal{F}$ into two disjoint sets $\mathcal{F}_\mathrm{MF}$ and $\mathcal{F}_\mathrm{BP}$ such that $\mathcal{F}_\mathrm{MF} \cup \mathcal{F}_\mathrm{BP} = \mathcal{F}$ and $\mathcal{F}_\mathrm{MF} \cap \mathcal{F}_\mathrm{BP} = \emptyset$. The MF (BP) subgraph contains the factor nodes in the set $\mathcal{F}_\mathrm{MF}(\mathcal{F}_\mathrm{BP})$ and the variable nodes connected to them. Here, we choose $\mathcal{F}_\mathrm{BP} = \cup_{t\in[T]}\{f_{\mathrm{u}_1^{(t)}}, ..., f_{\mathrm{u}_K^{(t)}}, f_{\mathrm{c}^{(t)}}, f_{\mathrm{m}_1^{(t)}}, ..., f_{\mathrm{m}_{N_\mathrm{d}}^{(t)}}\}$ and $\mathcal{F}_\mathrm{MF} = \mathcal{F}\backslash\mathcal{F}_\mathrm{BP}$.

### Support detection

We restrict the belief of the sparsity rate to be $q(\zeta) = \delta(\zeta - \hat{\zeta})$ where $\hat{\zeta}$ coincides with the mode of the unconstrained belief (see [11, Section II.D] for details), i.e.

$$\hat{\zeta} = \arg\max_{\zeta} \left\{ m_{f_\mathrm{a}\to\zeta}(\zeta) m_{f_\mathrm{s}\to\zeta}(\zeta) \right\} = \frac{\| \hat{\mathbf{s}} \|_0}{\bar{L}} \tag{C.5}$$

where $\hat{\mathbf{s}}$ is defined in the following. To that end we re-write (C.3) as

$$\mathbf{y} = \mathbf{A}\mathbf{s} + \mathrm{vec}(\boldsymbol{\Xi}), \tag{C.6}$$

where $\mathbf{A} \in \mathbb{C}^{NR \times \bar{L}}$ has the entries $A[a,l] = \sum_{r \in [R]} \Omega[a, (r-1)\bar{L} + l] v_l[r], a \in [NR], l \in [\bar{L}]$.

The belief of $\mathbf{s}$ reads $q(\mathbf{s}) \propto m_{f_\mathrm{o} \to \mathbf{s}}(\mathbf{s}) m_{f_\mathrm{s} \to \mathbf{s}}(\mathbf{s})$, where the two factors are $m_{f_\mathrm{s} \to \mathbf{s}}(\mathbf{s}) = \hat{\zeta}^{\|\mathbf{s}\|_0} (1-\hat{\zeta})^{\bar{L} - \|\mathbf{s}\|_0}$ and $m_{f_\mathrm{o} \to \mathbf{s}}(\mathbf{s}) = \mathrm{CN}(\mathbf{s} | \boldsymbol{\varepsilon}, \mathbf{Y})$, with $\boldsymbol{\varepsilon} = \mathbf{Y} \left\langle \lambda \mathbf{A}^\mathrm{H} \mathbf{y} \right\rangle_{\sim \mathbf{s}}$ and $\mathbf{Y}^{-1} = \left\langle \lambda \mathbf{A}^\mathrm{H} \mathbf{A} \right\rangle_{\sim \mathbf{s}}$. We define $\hat{\mathbf{s}} = \arg\max_{\mathbf{s} \in \{0,1\}^{\bar{L}}} \ln(q(\mathbf{s}))$, where

$$\ln(q(\mathbf{s})) = \left( \ln \frac{\hat{\zeta}}{1-\hat{\zeta}} \mathbf{1} + 2\Re\{ \left\langle \lambda \mathbf{A}^\mathrm{H} \mathbf{y} \right\rangle_{\sim \mathbf{s}} \} \right)^\mathrm{T} \mathbf{s} - \mathbf{s}^\mathrm{T} \mathbf{Y}^{-1} \mathbf{s} + \mathrm{const.} \tag{C.7}$$

where $\mathbf{1}$ is a vector with all $\bar{L}$ entries equal to one. To avoid an exhaustive search over $\mathbf{s}$, we use a single-replacement routine instead to update $\hat{\mathbf{s}}$ as described in Algorithm 1 (lines 6:9). The routine iteratively computes (C.7) for the vectors that differ from the current estimate $\mathbf{s}^\mathrm{old}$ at one location at most, i.e. with Hamming distance less than or equal to one. The vector that yields the maximum value of (C.7) is used to update the support.

### Channel weights estimation

The belief of the channel weight vector $\mathbf{v}_l, l \in [\bar{L}]$ reads $q(\mathbf{v}_l) \propto m_{f_\mathrm{o} \to \mathbf{v}_l}(\mathbf{v}_l) \cdot m_{f_{\mathrm{h}_l} \to \mathbf{v}_l}(\mathbf{v}_l) = \mathrm{CN}(\mathbf{v}_l | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$ with

$$\boldsymbol{\mu}_l = \boldsymbol{\Sigma}_l \left\langle \lambda \hat{s}[l] \left( {\boldsymbol{\Omega}^{(l)}}^\mathrm{H} \mathbf{y} - \sum_{k \neq l} \mathbf{J}^{(l,k)} \mathbf{v}_k^\mathrm{T} \right) \right\rangle_{\sim \mathbf{v}_l} \tag{C.8a}$$

$$\boldsymbol{\Sigma}_l = \left\langle \mathbf{K}^{-1} + \lambda |\hat{s}[l]|^2 \mathbf{J}^{(l,l)} \right\rangle_{\sim \mathbf{v}_l}^{-1} \tag{C.8b}$$

where we define $\mathbf{J}^{(l,k)} = \left( \boldsymbol{\Omega}^{(l)} \right)^\mathrm{H} \boldsymbol{\Omega}^{(k)}$ and $\boldsymbol{\Omega}^{(l)} = [\boldsymbol{\Omega}[\cdot, (r-1) \cdot \bar{L} + l], r \in [R]], l, k \in [\bar{L}]$.

### Noise precision estimation

The belief of the noise precision is $q(\lambda) \propto m_{f_\mathrm{n} \to \lambda}(\lambda) m_{f_\mathrm{o} \to \lambda}(\lambda)$, which yields the pdf of a Gamma distribution with mean

$$\langle \lambda \rangle_\lambda = NR \langle \| \mathbf{y} - \mathbf{A}\hat{\mathbf{s}} \|_2^2 \rangle_{\sim \lambda}^{-1}. \tag{C.9}$$

**Decoding**

The belief of $x_{\mathrm{d}}^{(t)}[k]$, $k \in [N_{\mathrm{d}}]$ is $q(x_{\mathrm{d}}^{(t)}[k]) \propto m_{f_{\mathrm{o}} \to x_{\mathrm{d}}^{(t)}[k]}(x_{\mathrm{d}}^{(t)}[k]) m_{f_{\mathrm{m}_k^{(t)}} \to x_{\mathrm{d}}^{(t)}[k]}(x_{\mathrm{d}}^{(t)}[k])$ where the first factor is $m_{f_{\mathrm{o}} \to x_{\mathrm{d}}^{(t)}[k]}(x_{\mathrm{d}}^{(t)}[k]) = \mathrm{CN}\left(x_{\mathrm{d}}^{(t)}[k] | m^{(t)}[k], v^{(t)}[k]\right)$ with

$$\begin{aligned} m^{(t)}[k] &= v^{(t)}[k] \left\langle \lambda(\mathbf{n}_k^{(t)})^{\mathrm{H}} \tilde{\mathbf{y}}_k^{(t)} \right\rangle_{\sim x_{\mathrm{d}}^{(t)}[k]} \\ v^{(t)}[k] &= \left\langle \lambda(\mathbf{n}_k^{(t)})^{\mathrm{H}} \mathbf{n}_k^{(t)} \right\rangle_{\sim x_{\mathrm{d}}^{(t)}[k]}^{-1}. \end{aligned} \tag{C.10}$$

The vectors $\mathbf{n}_k^{(t)} \in \mathbb{C}^R$ and $\tilde{\mathbf{y}}_k^{(t)} \in \mathbb{C}^R$ are defined as $\mathbf{n}_k^{(t)} = [M^{(t)}[(r-1)N + i_k, i_k], r \in [R]]^{\mathrm{T}}$, $\tilde{\mathbf{y}}_k^{(t)} = [\tilde{y}^{(t)}[(r-1)N + i_k], r \in [R]]^{\mathrm{T}}$, where $\mathbf{M}^{(t)} = [\mathrm{diag}(\mathbf{m}^{(1,t)}) ... \mathrm{diag}(\mathbf{m}^{(R,t)})]^{\mathrm{T}} \in \mathbb{C}^{NR \times N}$, $\mathbf{m}^{(r,t)} = \bar{\mathbf{D}}\bar{\mathbf{H}}[(t-1)L_\Delta + 1 : tL_\Delta, r] \in \mathbb{C}^N$ and $\tilde{\mathbf{y}}^{(t)} = \mathbf{y} - \sum_{t' \neq t} \mathbf{M}^{(t')} \mathbf{x}^{(t')}$. The second factor reads

$$m_{f_{\mathrm{m}_k^{(t)}} \to x_{\mathrm{d}}^{(t)}[k]}(x_{\mathrm{d}}^{(t)}[k]) \propto \sum_{\mathbf{c}_k^{(t)} \in \{0,1\}^Q} f_{\mathrm{m}_k^{(t)}}(x_{\mathrm{d}}{}^{(t)}[k], \mathbf{c}_k^{(t)}) \prod_{j=1}^{Q} n_{c_k^{(t)}[j] \to f_{\mathrm{m}_k^{(t)}}}(c_k^{(t)}[j]) \tag{C.11}$$

where $n_{c_k^{(t)}[j] \to f_{\mathrm{m}_k^{(t)}}}(c_k^{(t)}[j]) = m_{f_{\mathrm{c}^{(t)}} \to c_k^{(t)}[j]}(c_k^{(t)}[j])$. It is well known that propagating messages through node $f_{\mathrm{c}}^{(t)}$ corresponds to classical decoding e.g. [17], hence we do not detail the updates of $n_{c_k^{(t)}[j] \to f_{\mathrm{m}_k^{(t)}}}$.

**Algorithm summary and complexity study**

Algorithm 1 below summarizes the updates presented in Section 3.2. The initialization stage (lines 1:2) is followed by the iterative stage. Channel estimation (lines 4:11) and data detection and decoding (lines 12:16) are performed iteratively until a maximum number of iterations $M_{\mathrm{o}}$ has been reached.

We discuss next the main sources of complexity of the algorithm, namely a) the support detection (lines 6:9), b) the channel weights update (line 5) and c) the detection and decoding (lines 12:16).

**Support detection:** To update the support, (C.7) has to be computed for $\bar{L}+1$ vectors belonging to the set $\left\{\bar{\mathbf{s}} | d(\bar{\mathbf{s}}, \mathbf{s}^{\mathrm{old}}) \leq 1\right\}$. The computational complexity is dominated by the updates of the products $\left\langle \mathbf{A}^{\mathrm{H}} \right\rangle_{\mathbf{x}^{(1)},...,\mathbf{x}^{(T)},\mathbf{v}_1,...,\mathbf{v}_{\bar{L}}} \mathbf{y}$ and $\left\langle \mathbf{A}^{\mathrm{H}} \mathbf{A} \right\rangle_{\mathbf{x}^{(1)},...,\mathbf{x}^{(T)},\mathbf{v}_1,...,\mathbf{v}_{\bar{L}}}$ at the costs $\mathcal{O}(NR\bar{L})$ and $\mathcal{O}(NR\bar{L}^2)$ respectively.

**Channel weights updates:** To compute $\boldsymbol{\mu}_l$ and $\boldsymbol{\Sigma}_l$, $\left\langle \boldsymbol{\Omega}^{(l)} \right\rangle_{\mathbf{x}^{(1)},...,\mathbf{x}^{(T)}} \mathbf{y}$ and $\left\langle \left(\boldsymbol{\Omega}^{(l)}\right)^{\mathrm{H}} \boldsymbol{\Omega}^{(k)} \right\rangle_{\mathbf{x}^{(1)},...,\mathbf{x}^{(T)}}$, $l, k \in [\bar{L}]$ are updated at costs $\mathcal{O}(NR^2)$ and

$\mathcal{O}(NR^3)$ respectively. Note however that the matrices $\mathbf{\Omega}^{(l)}, l \in [\bar{L}]$ built by selecting $R$ columns of the matrix $\mathbf{\Omega} = \mathbf{I}_R \otimes \bar{\mathbf{\Phi}}$ are sparse, each holding $NR$ nonzero entries, hence the complexity can be further reduced to $\mathcal{O}(NR)$ for each of the two products above using a sparse BLAS routine [18]. The complexity of updating $\boldsymbol{\mu}_l$ and $\mathbf{\Sigma}_l$, $l \in [\bar{L}]$ are $\mathcal{O}(R^2)$ and $\mathcal{O}(R^3)$ respectively, the latter potentially being reduced to $\mathcal{O}(R^{2.373})$ [19].

**Detection and decoding:** We need to update the $TN_\mathrm{d}$ messages $m_{f_\mathrm{o} \to x_\mathrm{d}^{(t)}[k]}$, $k \in [N_\mathrm{d}]$ defined as the product of factors in (C.10). The cost of updating $m^{(t)}[k]$ and $v^{(t)}[k]$ is $\mathcal{O}(R)$ for each.

**Algorithm 1** Proposed iterative algorithm

set $\Delta\tau$, $L_\Delta$, $M_\mathrm{o}$, $M_\mathrm{i}$ and initialize $\langle\lambda\rangle_\lambda$, $\hat{\mathbf{s}}$
$\left\langle x_\mathrm{d}^{(t)}[k] \right\rangle_{x_\mathrm{d}^{(t)}[k]} = \left\langle |x_\mathrm{d}^{(t)}[k]|^2 \right\rangle_{x_\mathrm{d}^{(t)}[k]} = 0$, $k \in [N_\mathrm{d}]$, $t \in [T]$
**for** $o = 1 : M_\mathrm{o}$ **do**
  **for** $i = 1 : M_\mathrm{i}$ **do**
    compute $\mathbf{\Sigma}_l$ and $\boldsymbol{\mu}_l$ using (C.8) for $l \in [\bar{L}]$
    **do**
      $\mathbf{s}^\mathrm{old} = \hat{\mathbf{s}}$
      $\hat{\mathbf{s}} = \mathrm{argmax}_{\mathbf{s} \in \{\bar{\mathbf{s}} | d(\bar{\mathbf{s}}, \mathbf{s}^\mathrm{old}) \leq 1\}} \ln(q(\mathbf{s}))$ using (C.7)
    **while** $\hat{\mathbf{s}} \neq \mathbf{s}^\mathrm{old}$
    compute $\hat{\zeta}$ using (C.5) and $\langle\lambda\rangle_\lambda$ using (C.9)
  **end for**
  **for** $t \in [T]$ and $k \in [N_\mathrm{d}]$ **do**
    compute $m_{f_\mathrm{o} \to x_\mathrm{d}^{(t)}[k]}(x_\mathrm{d}^{(t)}[k])$ using (C.10)
    compute $m_{f_{\mathrm{m}_k^{(t)}} \to x_\mathrm{d}^{(t)}[k]}(x_\mathrm{d}^{(t)}[k])$ using (C.11)
    compute $\left\langle x_\mathrm{d}^{(t)}[k] \right\rangle_{x_\mathrm{d}^{(t)}[k]}$, $\left\langle |x_\mathrm{d}^{(t)}[k]|^2 \right\rangle_{x_\mathrm{d}^{(t)}[k]}$
  **end for**
**end for**

# 4 Numerical evaluation

We generate the MIMO system described in Section 2 using the parameter setting from Table C.2. The receiver is equipped with $R = 2$ antennas and receives signals from two transmitters, one serving, $(U = 1)$, and one interfering, $(I = 1)$. The pilots of both transmitters are uniformly spaced and have the same subcarrier pattern, i.e. the pilots of the transmitters overlap in frequency. We consider four variants of Algorithm 1 the features of which are detailed in the first four rows of Table C.3. NI neglects the interference. SM demodulates the signal of the interferer and estimates the impulse responses of its sub-channels based on the erroneous assumption that both transmitters use the same modulation alphabet. KM does the same operation as SM, but

in addition knows the modulation alphabet of the interferer. KMD knows the modulation and coding scheme (MCS) of the interferer and demodulates and decodes the interfering signal. It also estimates the responses of the interfering sub-channels. For benchmarking we use IF which implements Algorithm 1 and operates in interference-free conditions ($I = 0$), and Ref and RefD (see Table C.3) which implement [6] with a robust assumption on the prior covariance matrix [20]. We also use a genie-aided (GA) receiver possessing perfect CSI. Specifically, GA knows all $\mathbf{D}^{(t)}\mathbf{h}^{(t,r)}$ in (C.2) and the noise variance, and performs MMSE-IRC equalization [4]. GAIF is GA operating in interference-free conditions.

**Table C.2:** Parameter setting of the considered MIMO system

| **Transmitters** | | |
|---|---|---|
| | Serving transmitter | Interfering transmitter |
| N | 600 | 600 |
| pilot density | 25% | 25% |
| data modulation | 16-QAM | QPSK |
| pilot modulation | QPSK | QPSK |
| C | 1/3 | 1/3 |
| $d_{\mathrm{c}}$ | $4.69\mu s$ | $4.69\mu s$ |
| **Channel impulse response** | | |
| $p(L^{(t)}) = \mathrm{Poiss}(L^{(t)}\|5)$ and $p(\tau^{(t)}[l]) = \mathrm{U}(\tau^{(t)}[l]\|0, d_{\mathrm{c}})$ [2] | | |
| $p(h^{(t,r)}[l]\|\tau^{(t)}[l]) = \mathrm{CN}\left(h^{(t,r)}[l]\|0, V\exp(-10^6\tau^{(t)}[l])\right)$ [3] | | |
| where $t \in [T]$, $l \in [L^{(t)}]$, $r \in [R]$ and $V$ is a scaling parameter. | | |

The BER performance of the receivers in low and medium SIR regimes is shown in Fig. C.2 and C.3 respectively. IF performs very closely to GAIF. Its excellent performance indicates accurate channel and noise precision estimation, validating the usefulness of exploiting the SCS property. Obtaining the MCS of the interferer and decoding the interfering signal is highly beneficial in both SIR regimes and determines KMD to show a BER performance close to that of IF. This indicates good CI cancellation capabilities. Exploiting the SCS property proves to be advantageous for KMD which outperforms RefD. Not decoding the interfering signal causes KM and Ref to exhibit a BER degradation compared to KMD and RefD, however their performance is still similar to that of GA in both SIR regimes. The similar performance of KM and Ref stems from the inaccuracy in the detected symbols which does not allow for a further refinement of the channel estimates and therefore an accurate CI cancellation. At low SIR, even if SM performs worse than KM, it still outperforms NI. At medium SIR, SM exhibits less degradation compared to KM and Ref, revealing the benefit of reconstructing the signal and estimating the impulse responses of the interferer's sub-channels even under mismatched assumptions.

[2] The event $\{L^{(t)} = 0\}$ is discarded. Given $L^{(t)}$, the delays are drawn independently.
[3] The weights are drawn independently.

**Table C.3:** NI, SM, KM, KMD use the setting $M_\mathrm{o} = 10$, $M_\mathrm{i} = 5$, $\mathbf{K} = \mathbf{I}_R$, $\Delta\tau = 100ns$, $L_\Delta = \lceil \frac{d_\mathrm{c}}{\Delta\tau} \rceil$ and are initialized with $\langle\lambda\rangle_\lambda = 1$ and $\hat{\mathbf{s}} = \mathbf{1}$.

| Rx. | Knows MCS interferer | Demod. interfering signal | Decodes interfering signal | Estimates interferer channels |
|---|---|---|---|---|
| NI | – | – | – | – |
| SM | – | ✓ | – | ✓ |
| KM | ✓ | ✓ | – | ✓ |
| KMD | ✓ | ✓ | ✓ | ✓ |
| Ref | ✓ | ✓ | – | ✓ |
| RefD | ✓ | ✓ | ✓ | ✓ |

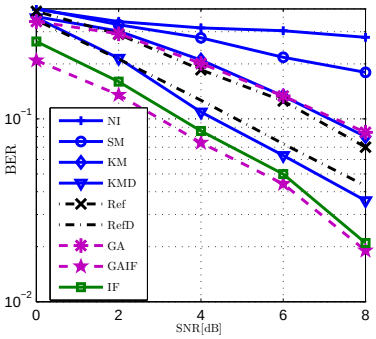


**Fig. C.2:** BER versus SNR at 0 dB SIR.

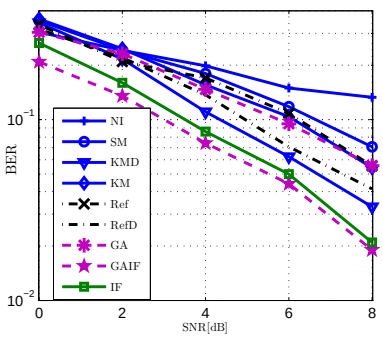


**Fig. C.3:** BER versus SNR at 5 dB SIR.

# 5 Conclusion

The numerical evaluation shows that a receiver implementing the proposed algorithm successfully cancels the co-channel interference and decodes the desired signal when it knows the interferer's modulation and code. Even when only the modulation alphabet is known, detecting the interfering signal and estimating the impulse responses of its sparse sub-channels is highly beneficial to the receiver. These results demonstrate the usefulness of exploiting the sparse common support property and acquiring information about the interferer's transmission. Even when the modulation alphabet of the interferer is unknown, a receiver that mistakenly assumes it to be identical to that of the desired data performs better than a receiver which neglects the interference entirely.

Lastly, we emphasize that the algorithm is capable of coping with an arbitrary number of interfering signals. Furthermore, its design allows for different scheduling options, therefore enabling the implementation of various interference-aware receivers that trade computational complexity for performance, and vice-versa.

## Acknowledgement

This work was supported by Intel Mobile Communications (IMC), the Danish Association for Science, Technology and Innovation, and the VIRTUOSO cooperative research project, funded by IMC, Anite, Telenor, Aalborg University and the Innovation Fund Denmark.

## References

[1] S. Wagner and F. Kaltenberger, "Interference-aware receiver design for MU-MIMO in LTE: Real-time performance measurements," *Intel Technology Journal, Volume 18, No1, 2014*.

[2] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 74–80, February 2014.

[3] P. Karunakaran, T. Wagner, A. Scherb, and W. Gerstacker, "On interference rejection combining for LTE-A systems: Analysis of covariance estimators and an iterative algorithm for frequency-selective channels," in *IEEE VTC*, May 2015, pp. 1–5.

[4] Y. Li and N. Sollenberger, "Adaptive antenna arrays for OFDM systems with cochannel interference," *Communications, IEEE Trans. on*, vol. 47, no. 2, Feb 1999.

[5] R. Ghaffar and R. Knopp, "Interference sensitivity for multiuser MIMO in LTE," in *IEEE SPAWC*, June 2011.

[6] C. Manchón, G. Kirkelund, B. Fleury, P. Mogensen, L. Deneire, T. Sorensen, and C. Rom, "Interference cancellation based on divergence minimization for MIMO-OFDM receivers," in *IEEE GLOBECOM*, Nov 2009.

[7] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, June 2010.

[8] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimating sparse MIMO channels having common support," in *IEEE ICASSP*, May 2011.

[9] M. Masood, L. Afify, and T. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *Signal Processing, IEEE Trans. on*, vol. 63, no. 1, Jan 2015.

[10] D. Wipf and B. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *Signal Processing, IEEE Trans. on*, vol. 55, no. 7, July 2007.

[11] E. Riegler, G. Kirkelund, C. Navarro, M. Badiu, and B. Fleury, "Merging belief propagation and the mean field approximation: A free energy approach," *IEEE Trans. Inf. Theory*, 2013.

[12] C. Manchón, G. Kirkelund, E. Riegler, L. P. Christensen, and B. Fleury, "Receiver architectures for MIMO-OFDM based on a combined VMP-SP algorithm," 2011.

[13] M.-A. Badiu, G. Kirkelund, C. Manchón, E. Riegler, and B. Fleury, "Message-passing algorithms for channel estimation and decoding using approximate inference," in *IEEE ISIT*, July 2012.

[14] C. Berger, S. Zhou, J. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," in *OCEANS - EUROPE*, May 2009.

[15] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., 1988.

[16] E. P. Xing, M. I. Jordan, and S. Russell, "A generalized mean field algorithm for variational inference in exponential families," in *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence*, ser. UAI'03, 2003.

[17] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb 2001.

[18] R. Pozo. (2006) Sparse basic linear algebra subprograms (BLAS) library. [Online]. Available: http://math.nist.gov/spblas/

[19] V. V. Williams, "Breaking the Coppersmith-Winograd Barrier," 2011. [Online]. Available: https://www.cs.rit.edu/~rlc/Courses/Algorithms/Papers/

[20] O. Edfors, M. Sandell, J.-J. van de Beek, S. Wilson, and P. Borjesson, "OFDM channel estimation by singular value decomposition," in *IEEE 46th VTC*, 1996.

# Paper D

## Sparse Channel Estimation Including the Impact of the Transceiver Filters with Application to OFDM

Oana-Elena Barbu, Niels Lovmand Pedersen, Carles Navarro Manchón, Guillaume Monghal, Christian Rom, Bernard H. Fleury

The paper has been published in the
*15th IEEE International Symposium on Signal Processing Advances in Wireless Communications (SPAWC)*
June 22 - June 25, Toronto, Canada, 2014.

*The layout has been revised.*

# Abstract

*Traditionally, the dictionary matrices used in sparse wireless channel estimation have been based on the discrete Fourier transform, following the assumption that the channel frequency response (CFR) can be approximated as a linear combination of a small number of multipath components, each one being contributed by a specific propagation path. In practical communication systems, however, the channel response experienced by the receiver includes additional effects to those induced by the propagation channel. This composite channel embodies, in particular, the impact of the transmit (shaping) and receive (demodulation) filters. Hence, the assumption of the CFR being sparse in the canonical Fourier dictionary may no longer hold. In this work, we derive a signal model and subsequently a novel dictionary matrix for sparse estimation that account for the impact of transceiver filters. Numerical results obtained in an OFDM transmission scenario demonstrate the superior accuracy of a sparse estimator that uses our proposed dictionary rather than the classical Fourier dictionary, and its robustness against a mismatch in the assumed transmit filter characteristics.*


# 1 Introduction

Many channel models proposed for wireless communication systems characterize the impulse response of the radio channel as the sum of a few dominant multipath components, each associated with a delay and a complex gain [1]. As a result, the channel frequency response (CFR), defined as the Fourier transform of the channel impulse response (CIR), admits a sparse representation in a specific Fourier dictionary. Methods from compressed sensing (CS) and sparse channel representations have been proposed to devise estimators of the radio channel responses that exploit this property [2–5].

However, the receiver of a wireless communication system observes a composite channel response that includes the impact of the propagation channel together with other effects, such as those induced by antenna or transceiver filter[1] responses. The combination of these effects results in a composite CFR that exhibits sparsity in a different (a priori unknown) dictionary. This naturally raises two questions: (i) can a CS-based estimator using classical Fourier dictionaries still yield precise sparse estimates of the composite channel?, and (ii) if this is not the case, which dictionary should the estimator use in order to produce accurate sparse estimates of the composite channel response?

The authors of [6] showed that the performance of CS algorithms is highly sensitive to mismatches in the used dictionary matrix. Such a situation is

[1] Hencefort, we use the term "transceiver filter" to designate either the transmit (shaping) or the receive (demodulation) filters.

encountered when CS-based channel estimators are applied to a signal model which incorrectly assumes perfect low-pass transceiver filters. To the authors' knowledge, only a few contributions have explored the effects of transceiver filters on sparse estimators before. In [3], the authors apply CS techniques to estimate the channel in a multicarrier system using perfect low-pass filters in highly mobile setups; they observe that the discrete delay-Doppler spreading function is approximately sparse. In this contribution we are, however, interested in the limitations that the practical implementation of such transceiver filters impose on the accuracy that sparse (or CS-based) estimators can attain. In [7], an OFDM system with transceiver filters is also analyzed. The author claims that, under the conditions of a sufficiently large bandwidth (e.g. 256 MHz), the resulting composite channel response appears approximately sparse. It is, however, unclear whether this conclusion holds for systems employing a smaller bandwidth as, e.g., an LTE system [1].

In this article we derive a model for the received OFDM signal with a dictionary explicitly accounting for the distortion introduced by transceiver filters. We then apply a CS-based channel estimator to this signal model and to the classical model which neglects this distortion [2, 4, 5]. Numerical investigations are conducted considering an LTE system as use-case. They reveal that the performance of the sparse estimator, measured in terms of mean squared error (MSE) of the CFR estimates, is significantly improved when it is applied in combination with our proposed dictionary, compared to when it is applied in combination with the classical Fourier dictionary. These investigations also demonstrate that the sparse estimator used in combination with the former dictionary is robust towards mismatches between the true and assumed characteristics of the filters, and that it performs well even in scenarios where the channel response exhibits a large number of multipath components.

The remainder of this paper is organized as follows. In Section 2 we derive an OFDM received signal model which includes the effects of the transceiver filters. Based on this, we propose in Section 3 a novel design of the dictionary used by sparse channel estimators. In Section 4 we test the performance of the aforementioned estimators. In Section 5 we sum up the observations and conclude the paper.

*Notation:* Boldface uppercase and lowercase letters designate matrices and vectors, respectively. The diagonal matrix $\mathbf{A} = \mathrm{diag}(\mathbf{a})$ has the entries of the vector $\mathbf{a}$ as diagonal elements. We denote by $[\mathbf{A}]_{i,j}$ the $(i,j)$th element of the matrix $\mathbf{A}$. We define the $N \times N$ discrete Fourier transform (DFT) matrix with $\mathbf{F} \in \mathbf{C}^{N\times N}, [\mathbf{F}]_{m,n} = 1/\sqrt{N}e^{-j2\pi mn/N}, m,n \in [0:N-1]$; $\mathbf{I}$ is the identity matrix. A function $f$ which maps the set $\mathcal{E}$ to the set $\mathcal{F}$ is denoted as $f : \mathcal{E} \to \mathcal{F}$ and its support is $\mathrm{supp}(f) = \{x \in \mathcal{E} |\ f(x) \neq 0\}$; the notation $|\mathcal{F}|$ denotes the cardinality of $\mathcal{F}$. We represent the convolution of two functions $f$ and $g$ as $f * g$; $\delta(\cdot)$ is the Dirac delta function. The notation $[P_1 : P_2]$ denotes the set $\{p \in \mathbb{N} | P_1 \leq p \leq P_2\}$. The superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transposition and

Hermitian transposition respectively. The notation $\| \mathbf{a} \|_{\mathbf{0}}$ denotes the number of non-zero entries of $\mathbf{a}$.

# 2 Signal model

We consider a single-input single-output (SISO) OFDM system model. By contrast to the traditional approach [5], we account for the response of the transceiver filters in the derivation of the model. The message consists of a vector $\mathbf{u} = [u_0, ..., u_{N_B-1}]^{\mathrm{T}}$ of information bits which are encoded with a code rate $R = N_B/N_C$ and interleaved to yield the vector $\mathbf{c} = [c_0, ..., c_{N_C-1}]^{\mathrm{T}}$. The code vector is modulated onto a vector of complex symbols that are multiplexed with the pilot symbols producing the symbol vector $\mathbf{x} = [x_0, ..., x_{N-1}]^{\mathrm{T}}$. The symbol $x_i$ is a pilot symbol if $i \in \mathcal{P}$ or a data symbol if $i \in \mathcal{D}$, where $\mathcal{P}$ and $\mathcal{D}$ represent the subsets of pilot and data indices respectively, so that $\mathcal{P} \cup \mathcal{D} = \{0, ..., N-1\}$, $\mathcal{P} \cap \mathcal{D} = \varnothing$, $|\mathcal{P}| = N_P$ and $|\mathcal{D}| = N_D$. We refer to $\mathcal{P}$ as the pilot pattern. The symbol vector $\mathbf{x}$ is passed through an inverse DFT block, yielding $\mathbf{s} = [s_0, ..., s_{N-1}]^{\mathrm{T}} = \mathbf{F}^{\mathrm{H}}\mathbf{x}$. Next, $\mathbf{s}$ is appended a $\mu$-sample long cyclic prefix (CP) and the entries of the resulting vector are modulated using a transmit shaping filter with impulse response $\psi_{\mathrm{tx}}(t)$ to produce the continuous-time OFDM signal

$$s(t) = \sum_{n=-\mu}^{N-1} s_n \psi_{\mathrm{tx}}(t - nT_s), \ t \in [-\mu T_s, NT_s) \tag{D.1}$$

where $T_s$ is the sampling period. We assume that $\mathrm{supp}(\psi_{\mathrm{tx}}) = [0, T]$, with $T = aT_s, a > 0$. The signal $s(t)$ is sent across a wireless channel with CIR $g(\tau)$ modeled as the sum of $L$ (specular) multipath components, with the complex gains $\boldsymbol{\beta} = [\beta_0, ..., \beta_{L-1}]^{\mathrm{T}}$ and delays $\boldsymbol{\tau} = [\tau_0, ..., \tau_{L-1}]^{\mathrm{T}}$:

$$g(\tau) = \sum_{l=0}^{L-1} \beta_l \delta(\tau - \tau_l). \tag{D.2}$$

We assume that $g(\tau)$ remains invariant over the duration of one OFDM symbol. At the reception, the signal appears as the convolution of the transmitted signal (D.1) and the CIR (D.2) corrupted by additive white Gaussian noise $n(t)$ with spectral height $\sigma^2$, i.e.

$$z(t) = (s * g)(t) + n(t). \tag{D.3}$$

The received signal is next passed through a receive demodulation filter with response $\psi_{\mathrm{rx}}(t)$,[2] $\mathrm{supp}(\psi_{\mathrm{rx}}) = [0, T]$, producing the output

$$r(t) = (z * \psi_{\mathrm{rx}})(t) = \sum_{n=-\mu}^{N-1} s_n (\psi_{\mathrm{tx}} * g * \psi_{\mathrm{rx}})(t - nT_s) + \nu(t) \tag{D.4}$$

[2]Without loss of generality, we assume $\psi_{\mathrm{rx}}(t)$ and $\psi_{\mathrm{tx}}(t)$ have energy one.

where $\nu(t) = (\psi_{\text{rx}} * n)(t)$. The output signal $r(t)$ is sampled and the CP is discarded, yielding the vector $\mathbf{r} = [r_0, ..., r_{N-1}]^{\text{T}}$ with the entries

$$r_k = r(kT_s) = \sum_{n=-\mu}^{N-1} s_n q((k-n)T_s) + \nu(kT_s), \tag{D.5}$$

$k \in [0 : N-1]$. In the above expression, we defined the composite channel response $q(t) = (g * \psi_{\text{tx}} * \psi_{\text{rx}})(t) = (g * \phi)(t), \text{supp}(q) \subseteq [0, \tau_{L-1} + 2T]$, with $\phi(t) = (\psi_{\text{tx}} * \psi_{\text{rx}})(t), \text{supp}(\phi) = [0, 2T]$. The noise samples in (D.5) form a circularly-symmetric complex Gaussian process with variance $\lambda^{-1}$, $\lambda \geq 0$ that is uncorrelated when the autocorrelation of $\psi_{\text{rx}}(t)$ satisfies the Nyquist criterion.[3]

We observe that decreasing the system bandwidth, i.e. increasing the sampling period $T_s$, results in widening the convolved response of the transceiver filters $\phi(t)$. As a result, for large bandwidths (small $T_s$) the composite response exhibits an approximately specular behavior as the response $\phi(t)$ decays fast, which justifies disregarding the filters' effects [7]. Conversely, when employing a smaller bandwidth (larger $T_s$) as e.g. in 20 MHz LTE systems, each multipath component in (D.2) is convolved with the slow-decaying response $\phi(t)$. Under such conditions, $q(t)$ is not well approximated by a specular response anymore.

In order to avoid inter-symbol interference, it must be ensured that $r_k = 0$ for $k > N + \mu$, which implies that $q((k-n)T_s) = 0$ for $k - n \geq \mu + 1$. When this condition is satisfied, the signal $\mathbf{y} = [y_0, ..., y_{N-1}]^{\text{T}}$ observed after the DFT processing at the receiver reads

$$\mathbf{y} = \mathbf{F}\mathbf{r} = \mathbf{X}\mathbf{M}\boldsymbol{\beta} + \boldsymbol{\xi} \tag{D.6}$$

where $\mathbf{X} = \text{diag}(x_0, ..., x_{N-1})$, $\mathbf{M} = \sqrt{N}\mathbf{F}\boldsymbol{\Phi}$, $\boldsymbol{\beta}$ is defined before (D.2), $\boldsymbol{\xi} = \mathbf{F}\boldsymbol{\nu}, \boldsymbol{\nu} = [\nu(0T_s), ..., \nu((N-1)T_s)]^{\text{T}} \in \mathbb{C}^N$, and $\boldsymbol{\Phi} \in \mathbb{R}^{N \times L}, [\boldsymbol{\Phi}]_{n,l} = \phi(nT_s - \tau_l), n \in [0 : N-1], l \in [0 : L-1]$.

In order to estimate the CFR at all subcarriers, i.e. $\mathbf{h} = \mathbf{M}\boldsymbol{\beta}$, we use the $N_P$ observations corresponding to the pilot subcarriers given by the pattern $\mathcal{P}$. The received signal observed at each pilot subcarrier $\mathbf{y}^{(\mathcal{P})}$ is divided by the corresponding known transmitted symbol. We note that the superindex $(\cdot)^{(\mathcal{P})}$ applied to a matrix $\mathbf{A}$ denotes a matrix $\mathbf{A}^{(\mathcal{P})}$ that contains the rows of $\mathbf{A}$ corresponding to the pattern $\mathcal{P}$. The vector of observations used for estimating the channel vector reads

$$\mathbf{t} = [\mathbf{X}^{(\mathcal{P})}]^{-1}\mathbf{y}^{(\mathcal{P})} = \mathbf{M}^{(\mathcal{P})}\boldsymbol{\beta} + [\mathbf{X}^{(\mathcal{P})}]^{-1}\boldsymbol{\xi}^{(\mathcal{P})}. \tag{D.7}$$

Thus, the observation $\mathbf{t}$ contains the samples of the CFR at the pilot subcarrier frequencies corrupted by noise. By contrast, the traditional observation model

[3] The receive filter's autocorrelation $A_{\text{rx}}(t) = \psi_{\text{rx}}(t) * \psi_{\text{rx}}(-t)$ satisfies the condition $A_{\text{rx}}(kT_s) = 0, \forall k \neq 0$.

[2, 5] disregards the effects of transceiver filters. In this case,

$$\mathbf{t} = \mathbf{T}^{(\mathcal{P})}\boldsymbol{\beta} + [\mathbf{X}^{(\mathcal{P})}]^{-1}\boldsymbol{\xi}^{(\mathcal{P})} \tag{D.8}$$

where $\mathbf{T} \in \mathbb{C}^{N\times L}$ has the entries $[\mathbf{T}]_{n,l} = e^{-j2\pi\frac{n}{NT_s}\tau_l}, n \in [0:N-1], l \in [0, L-1]$ and $\tau_l \in \boldsymbol{\tau}$. We will next discuss the effects of using the model (D.7) instead of (D.8) in the context of sparse channel estimation.

# 3 Compressed sensing inference for channel estimation

If the matrix $\mathbf{M}$ –and, hence, $\mathbf{M}^{(P)}$– was known, estimating $\mathbf{h}$ would be equivalent to estimating the entries of the vector of complex channel gains $\boldsymbol{\beta}$. Unfortunately, neither the dimensions of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ (i.e. the number of multipath components in (D.2)) nor the entries of $\boldsymbol{\tau}$, which the matrix $\mathbf{M}$ depends on, are known. In order to overcome this limitation we employ methods from CS to estimate the CIR and consequently the CFR $\mathbf{h}$. To that end, a discretized version of the CIR in (D.2) is used:

$$\bar{g}(\tau) = \sum_{k=0}^{K-1} \alpha_k \delta(\tau - \bar{\tau}_k) \tag{D.9}$$

where $\bar{\tau}_k = k\Delta_\tau, k \in [0:K-1]$ and we define $\boldsymbol{\alpha} = [\alpha_0, ..., \alpha_{K-1}]^{\mathrm{T}}$ and $\bar{\boldsymbol{\tau}} = [\bar{\tau}_0, ..., \bar{\tau}_{K-1}]^{\mathrm{T}}$.

## 3.1 Canonical compressed channel sensing model

Making use of (D.9), the canonical linear model used in compressed channel sensing [5, 8] approximates the observation model by:

$$\mathbf{t} \approx \mathbf{H}^{(\mathcal{P})}\boldsymbol{\alpha} + \mathbf{w} \tag{D.10}$$

where $\mathbf{w} \in \mathbb{C}^{N_P}$ is circularly-symmetric Gaussian distributed with zero-mean and covariance matrix $\lambda^{-1}\mathbf{I}$, and the matrix $\mathbf{H} \in \mathbb{C}^{N\times K}$ has entries

$$[\mathbf{H}]_{n,k} = e^{-j2\pi\frac{n}{NT_s}\bar{\tau}_k}, \tag{D.11}$$

$n \in [0:N-1], k \in [0:K-1]$. By choosing a sufficiently small sampling interval $\Delta_\tau$ and sufficiently large $K \gg L$, many entries of $\boldsymbol{\alpha}$ are expected to be either zero or close to zero, i.e. $\boldsymbol{\alpha}$ is expected to be approximately sparse.

Various CS methods [2, 5] have been proposed to compute sparse estimates of $\boldsymbol{\alpha}$ in (D.10). Once an estimate $\hat{\boldsymbol{\alpha}}$ has been obtained, the estimated CFR is computed as $\hat{\mathbf{h}} = \mathbf{H}\hat{\boldsymbol{\alpha}}$.

## 3.2 Novel compressed channel sensing model

We note that the model (D.10) is based on two drastic approximations of the true observation model (D.7): (i) the approximated observation model (D.8) that disregards the transceiver filters' effects, and (ii) the discretized CIR (D.9). While the approximation (ii) is necessary, we amend the approximation (i) and recast the CS model to

$$\mathbf{t} \approx \mathbf{H}_{\phi}^{(\mathcal{P})}\boldsymbol{\alpha} + \mathbf{w} \tag{D.12}$$

with $\mathbf{H}_{\phi} \in \mathbb{C}^{N\times K}$ defined as

$$[\mathbf{H}_{\phi}]_{n,k} = \sqrt{N}\sum_{m=0}^{N-1}[\mathbf{F}]_{n,m}\phi(mT_s - \bar{\tau}_k), \tag{D.13}$$

$n \in [0:N-1], k \in [0:K-1]$. We then apply CS techniques to the model (D.13) and obtain sparse estimates $\tilde{\boldsymbol{\alpha}}$; similarly, we compute the estimated CFR $\tilde{\mathbf{h}}_{\phi} = \mathbf{H}_{\phi}\tilde{\boldsymbol{\alpha}}$.

We note that the CFR $\mathbf{h}$ is $L$-sparse[4] in the dictionary $\mathbf{M}$ defined in (D.6). However, since $\mathbf{M}$ is unkown apriori, we employ the two approximate dictionaries $\mathbf{H}_{\phi}$ and $\mathbf{H}$ and consequently assume that $\mathbf{h}$ is approximately sparse in $\mathbf{H}_{\phi}$ and $\mathbf{H}$. In CS, the usage of approximate dictionaries is typically referred to as dictionary mismatch [6]. In the case of the dictionary $\mathbf{H}_{\phi}$, the mismatch is caused by the discretization of the delay domain carried out in (D.9); for the dictionary $\mathbf{H}$, an additional source of mismatch is present due to the neglection of the transceiver filter responses. Hence, we conjecture that a sparse estimator employing (D.12) instead of (D.10) will provide a more accurate estimate of $\mathbf{h}$. This conjecture is based on the fact that, when using the dictionary from (D.13), $\tilde{\boldsymbol{\alpha}}$ would correspond to an estimate of the CIR $g$ in (D.2), while utilizing (D.11) would yield an estimate $\hat{\boldsymbol{\alpha}}$ of the composite channel $q$ defined after (D.5). For small-to-medium bandwidths, the latter approach will result in estimates of $\boldsymbol{\alpha}$ with more non-zero entries than those obtained with the former approach, due to the effect of the filters response $\phi$ on the composite response $q$. Thus, a less accurate reconstruction of the CFR is expected.

## 3.3 Sparse channel estimation using sparse Bayesian learning

We use a Bayesian inference method commonly referred to as sparse Bayesian learning (SBL) to obtain sparse estimates of $\boldsymbol{\alpha}$ in the CS models in (D.10) and (D.12). SBL makes use of models of the prior pdf $p(\boldsymbol{\alpha})$ that strongly penalize non-sparse estimates in maximum-a-posteriori based estimators [9], [10]. In

[4] A signal $\mathbf{a}$ is $L$-sparse in a dictionary $\mathbf{A}$ if a vector $\mathbf{b}$ exists with $\| \mathbf{b} \|_0 = L$ s.t. $\mathbf{a} = \mathbf{A}\mathbf{b}$.

**Table D.1:** Parameter settings

| | |
|---|---|
| Sampling time $T_s$ | 32.55 ns |
| Bandwidth $B$ | 20 MHz |
| CP length | 144 $T_s$ |
| Modulation | 64 QAM |
| Turbo-encoder rate | 948/1024 |
| Decoder | BCJR [12] |
| Number of subcarriers $N$ | 1200 |
| Number of pilots/time slot $N_P$ | 400 |

this work, we have selected the prior model proposed in [2], where the prior pdf of $\boldsymbol{\alpha}$ is formulated as

$$p(\boldsymbol{\alpha}) = p(\boldsymbol{\alpha}; \epsilon, \eta) = \int_0^\infty p(\boldsymbol{\alpha}|\boldsymbol{\gamma}) p(\boldsymbol{\gamma}; \epsilon, \eta) \mathrm{d}\boldsymbol{\gamma} \tag{D.14}$$

with $p(\boldsymbol{\alpha}|\boldsymbol{\gamma}) = \prod_{k=0}^{K-1} p(\alpha_k|\gamma_k)$, $p(\boldsymbol{\gamma}; \epsilon, \eta) = \prod_{k=0}^{K-1} p(\gamma_k; \epsilon, \eta)$, where $p(\alpha_k|\gamma_k)$ is a Gaussian pdf with zero-mean and variance $\gamma_k$, and $p(\gamma_k; \epsilon, \eta)$ is a Gamma pdf with shape and rate parameters $\epsilon$ and $\eta$, respectively. Using the above prior model, the estimation algorithm presented in [2] is applied to the observation models (D.10) and (D.12).

# 4 Performance evaluation

## 4.1 Setup

In this section we study the performance of the SBL channel estimator using our proposed dictionary matrix design in a SISO LTE OFDM setup [1], with the settings specified in Table D.1. The pilots are arranged according to the pattern specified in [11]. Both transmit and receive filters are truncated square-root raised cosine filters with roll-offs $r_{\mathrm{TX}}$ and $r_{\mathrm{RX}}$ respectively, and duration $T = 3T_s$.

We employ two different versions of the SBL estimator proposed in [2]. One version of the estimator, which we coin SE, uses the classical dictionary matrix design in (D.11). The second version, referred to as SE(F) henceforth, uses our proposed dictionary matrix design in (D.13), which accounts for the responses of the transceiver filters. For both estimators, we use the parameter setting $(\epsilon = 0.5, \eta = 1, K = 500, \Delta_\tau = 10\mathrm{ns})$, see (D.9) and (D.14). For benchmarking purposes, we consider two additional estimators: (i) a genie-aided estimator (GAE) [13] that assumes perfect knowledge of multipath components delays, i.e. uses the dictionary $\mathbf{M}$ from (D.6), and (ii) a robust design of the classical Wiener filter estimator, which we refer to as robust Wiener filter (RWF). The

**Table D.2:** Scenario A. Channel Power Delay Profile

| Delays [$\mu$s] | 0 | 0.5 | 1.6 | 2.3 | 3.3 |
|---|---|---|---|---|---|
| Power [dB] | -1 | 0 | -3 | -5 | -7 |

latter estimator is designed assuming that the channel has a maximum excess delay of 5 $\mu$s and a robust covariance matrix following [14]. Note that this assumption is also implicitly made in the design of SE and SE(F) via the aforementioned parameter setting.

We assess the accuracy of the investigated estimators by evaluating the performance, in terms of MSE of the CFR estimates and coded bit-error rate (BER), of a receiver employing them. The receiver is evaluated in two different propagation scenarios, each characterized by a specific channel model. In both scenarios, block fading is assumed. Scenario A employs a sparse 3GPP-like channel, modeled as specified in [1]: the CIR consists of five multipath components, with associated delays drawn independently from a uniform distribution with a 10 ns range, centered around the delays specified in Table D.2.[5] Scenario B employs the model introduced in [15], and later studied in [16]. The CIRs generated from the model exhibit a number of clustered multipath components which varies over different realizations, with cluster and within-cluster delays following Poisson arrival processes with rates $\Lambda$ and $\lambda$, respectively. The conditional second moments of the channel gains are modeled by the power-delay constants $(\Gamma, \gamma)$. We set $[1/\Lambda, 1/\lambda]$ $[\mu\text{s}] = [0.3, 5]$ and $[\Gamma, \gamma]$ [ns] $= [600, 200]$, leading to channel realizations that contain, in average, fifteen multipath components. Scenario A enables us to determine whether the filters' effects impair the receiver performance when they are not accounted for. Scenario B allows the total number of multipath components to vary over different realizations in order to account for the variabiliy of scatterers in the environment, introducing therefore an additional degree of freedom compared with the channel in Scenario A. The harsher channel conditions of Scenario B allow us to draw further conclusions on the performance of the studied estimators.

## 4.2 Numerical results

In Fig. D.1 we depict the performance of the sparse estimators in terms of MSE of the CFR in Scenarios A and B respectively when the transmit and receive filters are perfectly matched ($r_{\text{TX}} = r_{\text{RX}} = 0.5$). In both scenarios, we observe that the mismatched dictionary matrix used in SE degrades the estimator's performance as the SNR increases. Since it accounts for the filter's responses in the dictionary matrix, SE(F) performs closely to GAE for all SNR values;

[5] We ensure in this way that the true delays are not integer multiples of the delay resolution we have selected for the sparse estimator - see (D.9).

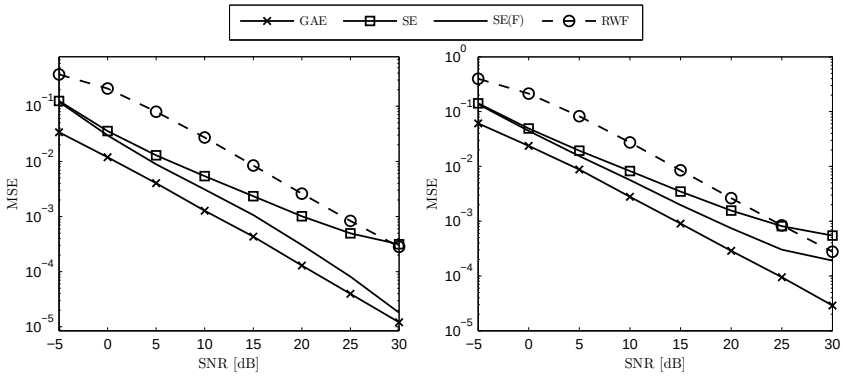


**Fig. D.1:** Comparison of the MSE performance of the investigated channel estimators in Scenario A (left) and Scenario B (right).

the slight performance degradation that SE(F) suffers in Scenario B at high SNR is caused by the presence of numerous multipath components. However, in both scenarios SE(F) outperforms SE and RWF.

In practical situations, the characteristics of the transmitter's radio frequency front end may be unknown by the receiver. As a result, the receiver often possesses incomplete information for computing the dictionary matrix. This provides an incentive to study how the mismatch between the transmit and receive filters affects the accuracy of the estimator. To conduct this investigation, we fix the receive filter roll-off to $r_{\mathrm{RX}} = 0.5$, and vary the transmit filter roll-off $r_{\mathrm{TX}}$. The GAE and SE(F) estimators assume $r_{\mathrm{TX}} = r_{\mathrm{RX}}$, regardless of the actual value of $r_{\mathrm{TX}}$. The resulting MSE is depicted in Fig. D.2 for the two scenarios at 30 dB SNR. As expected, SE(F) achieves its best performance when the roll-off factor of the transmit and receive filters coincide ($r_{\mathrm{TX}} = r_{\mathrm{RX}} = 0.5$). However, even in the case of a roll-off mismatch, SE(F) always outperforms SE. Hence, SE(F) is robust against mismatches of the assumed filter roll-off parameters.

Fig. D.3 depicts the BER performance of a receiver employing the investigated channel estimators. We have selected as an ideal reference a receiver which has knowledge of the true CFR coefficients. We observe that, in Scenario A, SE(F) performs almost as well as the ideal reference, with a gain of up to 1 dB with respect to SE. The harsher channel conditions in Scenario B lead to a less significant benefit of using SE(F). Nonetheless, a receiver employing SE(F) always performs better than a receiver employing SE.

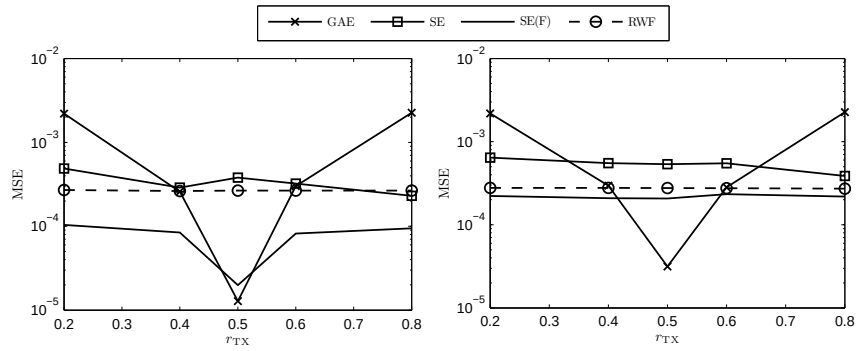


**Fig. D.2:** Comparison of the robustness of the investigated estimators towards mismatched filter parameters ($r_{\mathrm{RX}} = 0.5,\ r_{\mathrm{TX}} \in [0.2\ 0.8]$) in Scenario A (left) and Scenario B (right). The estimators assume $r_{\mathrm{TX}} = r_{\mathrm{RX}} = 0.5$.

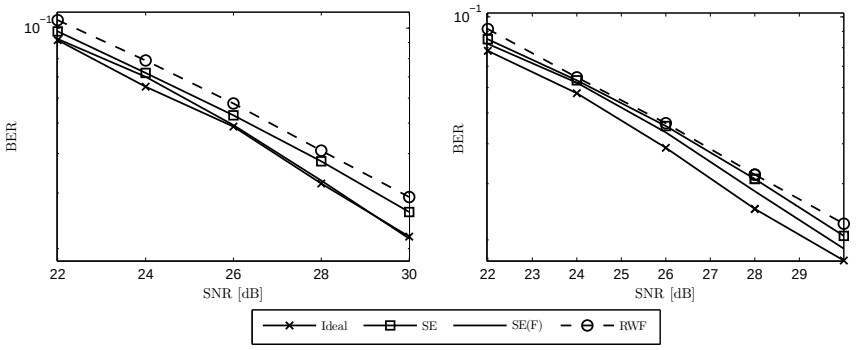


**Fig. D.3:** Comparison of the BER performance of a receiver using SE, SE(F), and RWF in Scenario A (left) and Scenario B (right).

# 5 Conclusion

In this paper we have analyzed the effect that transceiver filters have on the accuracy of selected state-of-art sparse channel estimation techniques. Traditional CS techniques for channel estimation employ Fourier dictionaries, which fail to embed the transceiver filters' responses. As a result, the CS-based channel estimators operate with mismatched dictionaries which degrade their estimation accuracy. To overcome this limitation, we have proposed a novel design of the dictionary matrix which accounts for the filters' responses, allowing thus for sparser representations of the channel response.

To evaluate the validity of the proposed solution, we applied an SBL estimator that includes either this new dictionary or the traditional Fourier dictionary

to an OFDM communication system. Numerical results illustrated that the SBL estimator employing our dictionary design always performs better than when it uses the classical dictionary. Additionally, we observed that our proposed dictionary matrix is especially advantageous in scenarios in which the channel exhibits a high degree of sparsity. Furthermore, even when the receiver possesses imperfect information about the filters' responses, the proposed dictionary yields a robust behavior of the sparse estimator.

Finally, we point out that, even though this study has been restricted to a particular choice of estimator, the proposed dictionary can be applied to any sparse channel estimator derived within the CS framework. Hence, we conclude that the dictionary matrix design proposed in this work will be a valuable tool to enable robust and accurate CS-based channel estimation in future generations of wireless receivers.

## Acknowledgment

This work was supported in part by the 4GMCT cooperative research project funded by Intel Mobile Communications, Agilent Technologies, Aalborg University and the Danish National Advanced Technology Foundation.

## References

[1] "Evolved Universal Terrestrial Radio Access; LTE Physical Layer - Base Station radio transmission and reception (Rel. 8) 2012," Tech. Rep.

[2] N. L. Pedersen, C. Navarro Manchón, and B. H. Fleury, "A fast iterative Bayesian inference algorithm for sparse channel estimation," in *IEEE Int. Conf. Comm.*, 2013.

[3] G. Tauböck and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," in *ICASSP*, 2008, pp. 2885–2888.

[4] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. of the IEEE*, vol. 98, no. 6, 2010.

[5] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, 2010.

[6] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, 2011.

[7] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, 2011.

[8] N. L. Pedersen, C. N. Manchón, D. Shutin, and B. H. Fleury, "Application of Bayesian hierarchical prior modeling to sparse channel estimation," in *IEEE Int. Conf. Comm.* IEEE, 2012, pp. 3487–3492.

[9] M. E. Tipping, "Sparse Bayesian learning and the Relevance Vector Machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[10] M. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, 2003.

[11] "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8)," Tech. Rep., 2009.

[12] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 284–287, 1974.

[13] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, 2001.

[14] O. Edfors, M. Sandell, J.-J. Van de Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, 1998.

[15] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *Selected Areas in Communications, IEEE Journal on*, vol. 5, no. 2, pp. 128–137, February 1987.

[16] M. L. Jakobsen, T. Pedersen, and B. H. Fleury, "Analysis of the stochastic channel model by Saleh & Valenzuela via the theory of point processes," in *Int. Zurich Seminar on Comm.*, 2012, pp. 115–118.

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-542-9

AALBORG UNIVERSITY PRESS